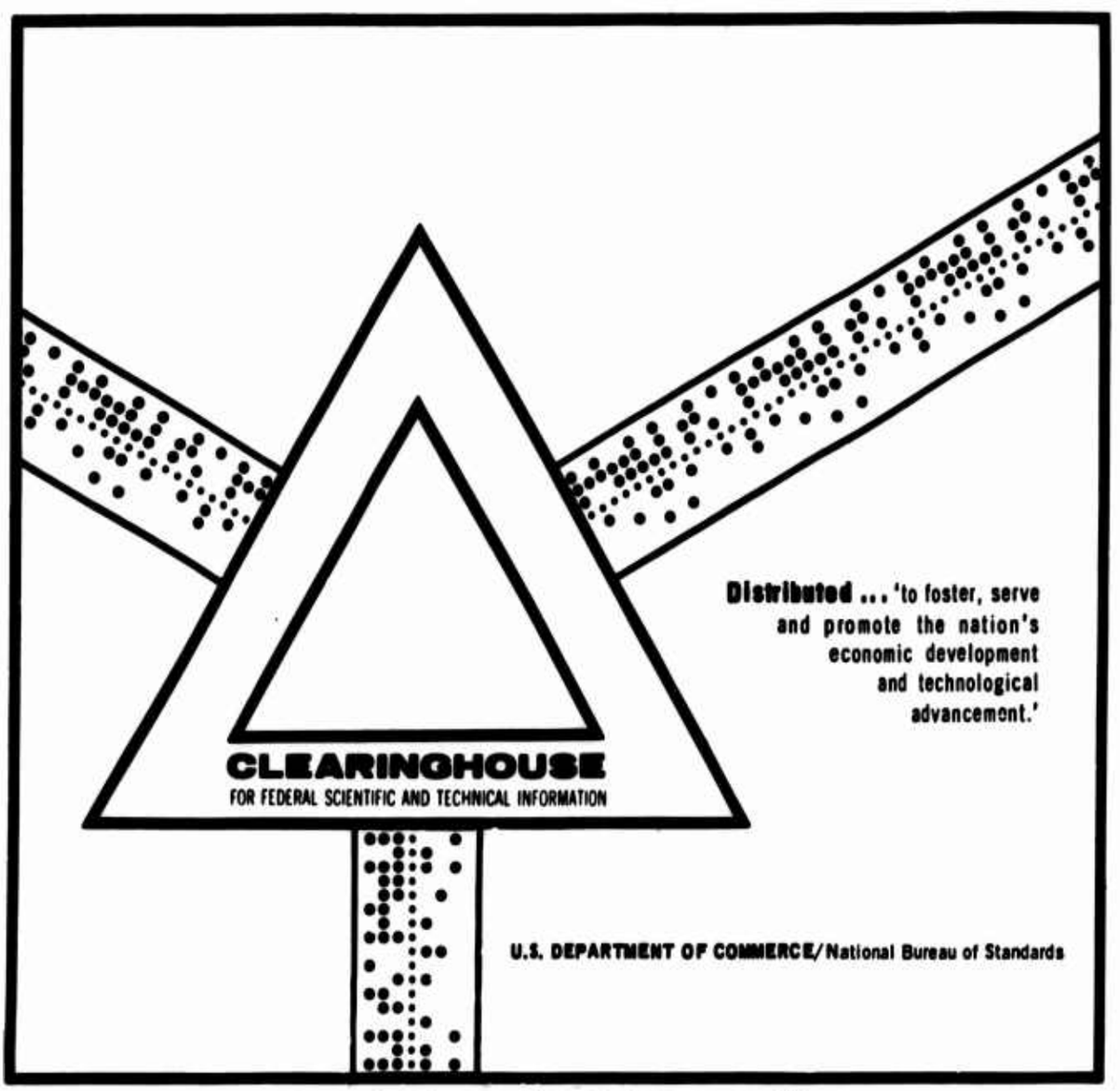
DEVELOPMENT OF NEW METHODS FOR THE SOLUTION OF DIFFERENTIAL EQUATIONS BY THE METHOD OF LIE SERIES

W. Groebner, et al

Innsbruck University
Innsbruck, Austria

July 1969



DEVELOPMENT OF NEW METHODS FOR THE SOLUTION OF DIFFERENTIAL EQUATIONS BY THE METHOD OF LIE SERIES

Final Technical Report

By

W. Gröbner, K.H. Kastlunger, H. Reitberger, R. Sály, G. Wanner

July 1969

EUROPEAN RESEARCH OFFICE

Contract No. JA 37-68-C-1199

Contractor: Prof. W. Gröbner
Department of Mathematics
University of Innsbruck
Austria

DEVELOPMENT OF NEW METHODS FOR THE SOLUTION OF DIFFERENTIAL EQUATIONS BY THE METHOD OF LIE SERIES

Final Technical Report

By

W. Gröbner, K.H. Kastlunger, H. Reitberger, R. Säly, G. Wanner

July 1969

EUROPEAN RESEARCH OFFICE

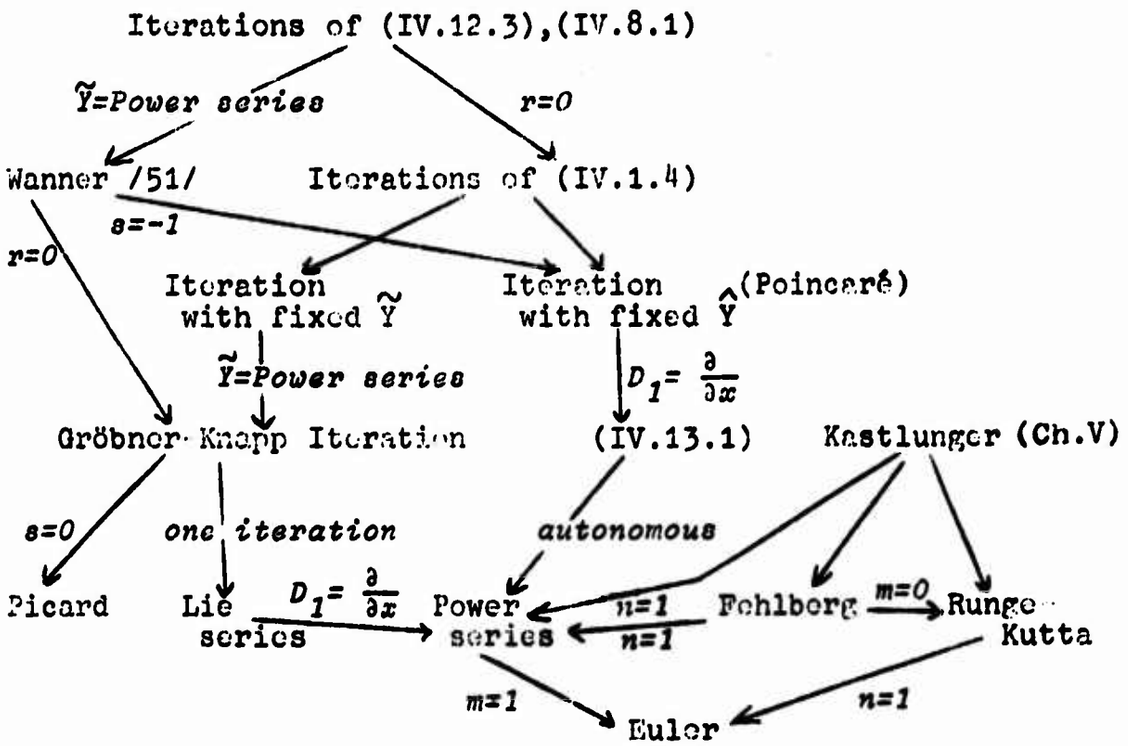
Contract No. JA 37-68-C-1199

Contractor: Prof. W. Gröbner
Department of Mathematics
University of Innsbruck
Austria

Summary

This report summarizes the recent work in the application of the LIE-series method to the solution of ordinary and partial differential equations.
After a short introduction the power series method which is a special case of the Lie series method of chapter III is described in chapter II. Further we discuss the interesting concept of recursion formulas and the calculation of the "transfer matrix" (connection matrix), the derivatives of the solution with respect to the initial values.
Chapter III deals with the numerical evaluation of the Lie series perturbation formula. This chapter contains the results of the report /29/, which has been written together with H. Knapp at the MRC, Madison, Wisconsin. Suitable quadrature formulas and recursions, statements on the order and error estimation are given. Numerical examples finish the chapter and compare the method also with that of Fehlberg.
In chapter IV we prove Gröbner's integral equation which leads to short proofs of the formulas of chapter III and to various generalizations of the method. A survey of these is presented at the end of this summary.
Chapter V generalizes the concept of Runge-Kutta to methods with multiple nodes, which is possible with the use of the Lie differential operator D. A general theory is developed and the method of Fehlberg is shown to be a special case.
Chapter VI deals with the step-size control and chapter VII shows the application of generalized Lie series to the calculation of switch-on transients occuring in the telegraphic equation.

SURVEY ON THE METHODS OF THE REPORT



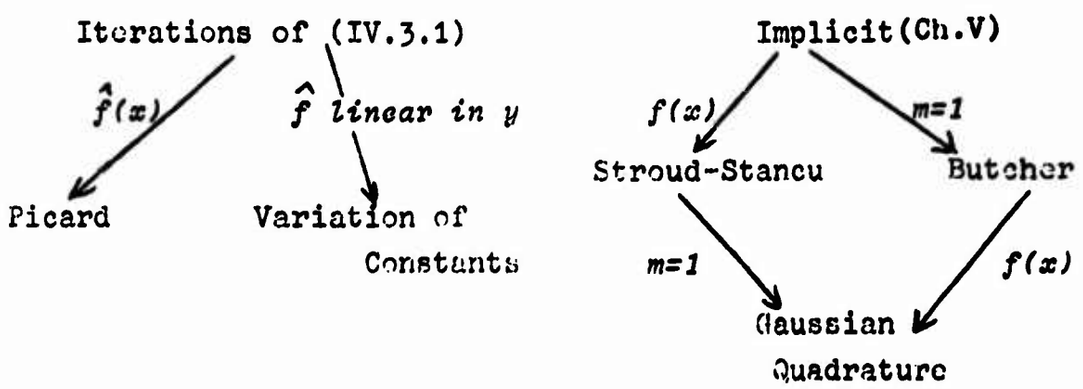


TABLE OF CONTENTS

Chapter I: INTRODUCTION

I.1. Statement of the problem

Find the solutions $y_1(x),\dots,y_n(x)$ of an ordinary system of first-order differential equations

$$\begin{aligned} y_1' &= f_1(x,y_1,\dots,y_n) \\ &\dots\dots\dots\dots\dots \\ y_n' &= f_n(x,y_1,\dots,y_n) \end{aligned} \tag{1.1}$$

which at x_o assume n specified initial values

$$y_i(x_o) = y_{io} \qquad (i=1,\dots,n) \tag{1.2}$$

Here, $f_i(x,y_1,\dots,y_n)$ are given functions of the variables $x,y_1,\dots,y_n$. Defining the vectors

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

we write (1.1) as

$$y' = f(x,y) \tag{1.3}$$

We shall keep to this way of writing in what follows. Speaking, for example, of "the solution y(x)" we mean that this ist the solution vector

$$y(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix}$$

i.e., "the solutions $y_1(x),\dots,y_n(x)$" , etc.

When stated as above, our problem is already quite general because any explicit higher-order differential equation or system can be re-written as a first-order system. This requires merely that all derivatives except the highest be replaced by new auxiliary functions (cf. Erwe /11/, p. 27).

I.2. Step-by-Step Continuation of Solutions

All methods discussed in the following give reliable approximations

$\hat{y}(x)$ only in the near neighborhood of the initial value x_0. Large values of $|x-x_0|$ may soon lead to poor results. What one can do is choose a certain "step" h_1 and trace the approximation only to the point $x_1=x_0+h_1$. This approximation $\hat{y}(x_0+h)$ will then serve as the initial value of a new step from x_1 to $x_2=x_1+h_2$, and so forth.

Apart from the specified initial value, such "one-step-methods" do not use any other information on the previous shape of the solution. Therefore, we need no longer bother to number the steps but may call any initial point x_0,y_0. The problem left for the following chapters is now to construct an approximation $\hat{y}(x)$ at the point $x=x_0+h$ from given initial values x_0,y_0 and a given step size h with a sensible volume of calculation in such a way that this approximation is as close as possible to the unknown solution.

I.3. Error

The size h of the steps depends above all on the desired accuracy. Smaller steps give better accuracy (not considering rounding errors) but require more work. To make a sensible choice of the step size we must therefore have a rough idea of the "local" error committed during a step of integration. We shall discuss this when dealing with the different methods individually. However, the total error committed after several steps is still undetermined. This error may soon become much greater than would be expected because of the insignificant local errors. The decisive factor is whether the solutions next to $y(x)$ approach $y(x)$ or depart from it as x increases, i.e. whether the solution is stable or unstable. More information about this can be got from the so-called transfer matrix. In Section II.6 we will see how to calculate it.

In the case of n=1, i.e., one differential equation, only half of all cases give unstable solutions. In systems of differential equations (hence, also in differential equations of higher order), however, there is nearly always at least one unstable component. Therefore, accuracy must be high should the solution be continued over a domain of considerable extent. Here are two examples that involve some trouble:

$$y''=10y'+11y, \quad y(0)=1, \quad y'(0)=-1$$
$$y(3)=?$$

(For greater detail see Collatz /7/, p. 49),

$$y''+(1-x^2)y=0, \quad y(0)=1, \quad y'(0)=0, \quad y(100)=?.$$

In the last example, accuracy would have to be 5000 places if something should be obtained for x=100.

Chapter II

Power Series

by G. Wanner

Abstract:

Solving ordinary differential equations by power series expansions has again become rather popular lately, on the one hand because the coefficients of the solutions can be calculated by computer through recursion formulas, and on the other hand because estimation of error is relatively simple.

BLANK PAGE

II.1. Solution by Power Series Expansion

Power series of the solutions $y(x)$ will henceforth be written in the way adopted by W. Groebner. This will turn out to be very useful, especially in later chapters.

Let $F(x,y)$ be an analytic function of the variables $x,y_1,\ldots,y_n$. Inserting solutions $y_1(x),\ldots,y_n(x)$ in the place of $y_1,\ldots,y_n$ we find a function that depends on x only. By the chain rule, its derivative with respect to this variable is

$$(1.1)\qquad \frac{d}{dx}F(x,y(x)) = \left[F_x+F_{y_1}\frac{dy_1}{dx} + \ldots +F_{y_n}\frac{dy_n}{dx}\right]_{x,y(x)}$$

Here, the bracket symbol $\left[\ \ \right]_{x,y(x)}$ means that the variables x and y must be replaced by the functions x and $y(x)$ <u>after</u> the partial differentiations have been performed. From now on we shall keep to this way of writing, i.e., every time some kind of expression stands after such brackets it must be inserted for the variables x and y. Since the functions $y(x)$, which we have inserted in Eq. (1.1), are supposed to be the solutions of (I.1.1) or (I.1.3) we have

$$\frac{dy_i}{dx} = f_i(x,y(x)) = \left[f_i(x,y)\right]_{x,y(x)}$$

and

$$\frac{d}{dx}F(x,y(x)) = \left[\frac{\partial F}{\partial x}+f_1(x,y)\frac{\partial F}{\partial y_1}+\ldots+f_n(x,y)\frac{\partial F}{\partial y_n}\right]_{x,y(x)} =$$

$$(1.2)\qquad = \left[DF\right]_{x,y(x)}$$

where we have defined the linear differential operator

$$(1.3)\qquad D = \frac{\partial}{\partial x}+f_1(x,y)\frac{\partial}{\partial y_1}+\ldots+f_n(x,y)\frac{\partial}{\partial y_n}$$

for brevity.

By iteration of (1.3) we find for the higher derivatives

$$(1.2')\qquad \frac{d^\mu}{dx^\mu}F(x,y(x)) = \left[D^\mu F\right]_{x,y(x)},$$

where $D^\mu F$ means that the differential operator has to be applied μ times to F.

Thus, the power series of the functions $F(x,y(x))$ at the point x_0

can be written as

$$(1.4)\qquad F(x,y(x)) = \sum_{\mu=0}^{\infty} \frac{(x-x_o)^{\mu}}{\mu!} \frac{d^{\mu}}{dx^{\mu}} \Big[F(x,y(x)) \Big]_{x=x_o} =$$

$$= \sum_{\mu=0}^{\infty} \frac{(x-x_o)^{\mu}}{\mu!} \Big[D^{\mu} F \Big]_{x_o,y_o}$$

owing to $y(x_o)=y_o$ (I.1.2).

Setting $F(x,y)=y_i$ we obtain the series for the solutions proper

$$(1.5)\qquad y_i(x) = \sum_{\mu=0}^{\infty} \frac{(x-x_o)^{\mu}}{\mu!} \Big[D^{\mu} y_i \Big]_{x_o,y_o} \qquad (i=1,\ldots,n).$$

Similar expressions occur also in the theory of transformation groups. Therefore, such series, especially the ones derived in the following chapters, are also named Lie-series.

II.2. Recursive Calculation of the Coefficients

We shall now discuss the recursive calculation of the power series coefficients as lately adopted by Gibbons /18/, R.E. Moore /36/ and many other authors. It has become very important thrcugh the use of electronic computers.

We assume that the functions $f_1(x,y)$ have been composed of the variables x and $y_1,\ldots,y_n$ by finite sequences of elementary operations. We note all intermediate results. Each of these intermediate results follows from one or two of the preceding values (one-place and/or binary operations) or from $x,y_1,\ldots,y_n$ or from a constant c. Suppose $p(x,y_1,\ldots,y_n)$, $q(x,y_1,\ldots,y_n)$ and $r(x,y_1,\ldots,y_n)$ are three (or two) operands that are interrelated through an arithmetic operation

$$(2.1)\qquad r(x,y_1,\ldots,y_n) = p(x,y_1,\ldots,y_n) * q(x,y_1,\ldots,y_n)$$

or some sort of elementary functions g

$$(2.2)\qquad r(x,y_1,\ldots,y_n) = g(p(x,y_1,\ldots,y_n)) \quad .$$

Then we introduce the following notation

(2.3) $P_\mu = \frac{D^\mu p}{\mu!}$, $Q_\mu = \frac{D^\mu q}{\mu!}$, $R_\mu = \frac{D^\mu r}{\mu!}$.

Hence, these quantities are functions of the variables $x, y_1, \ldots, y_n$. For the functions $x, y_1, \ldots, y_n, f_1, \ldots, f_n, c$ which are special cases of such operands, we shall also use the corresponding symbols X_μ, $Y_{1\mu}, \ldots, Y_{n\mu}$, $F_{1\mu}, \ldots, F_{n\mu}$, C_μ .

In what follows we tabulate formulas which permit us to calculate R_μ for (2.1) or (2.2) from the coefficients

$$P_\mu , P_{\mu-1} , P_{\mu-2} , \ldots , P_0$$
$$Q_\mu , Q_{\mu-1} , Q_{\mu-2} , \ldots , Q_0 \qquad \text{(only for (2.1))}$$
$$R_{\mu-1} , R_{\mu-2} , \ldots , R_0 \ .$$

Sum: $r=p+q$	$R_\mu = P_\mu + Q_\mu$
Difference: $r=p-q$	$R_\mu = P_\mu - Q_\mu$
Product: $r=p.q$	$R_\mu = \sum_{\rho=0}^{\mu} P_\rho Q_{\mu-\rho} \quad (\mu=0,1,2,\ldots)$
Quotient: $r=p/q$	$R_\mu = (P_\mu - \sum_{\rho=0}^{\mu-1} R_\rho Q_{\mu-\rho})/Q_0 \quad (\mu=0,1,2,\ldots)$
exp: $r=\exp p$	$R_\mu = \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)R_\rho P_{\mu-\rho}$, $R_0=\exp P_0$
log: $r=\log p$	$R_\mu = \{P_\mu - \frac{1}{\mu}\sum_{\rho=1}^{\mu-1}(\mu-\rho)P_\rho R_{\mu-\rho}\}/P_0$, $R_0=\log P_0$
Square root: $r=\sqrt{p}$	$R_\mu = \frac{1}{2R_0}\{P_\mu - \sum_{\rho=1}^{\mu-1} R_\rho R_{\mu-\rho}\}$, $R_0=\sqrt{P_0}$
Constant power. $r=p^c$	$R_\mu = \{\sum_{\rho=0}^{\mu-1}(c\mu-(c+1)\rho)R_\rho P_{\mu-\rho}\}\frac{1}{\mu P_0}$, $R_0=P_0^c$ ($P_0 \neq 0$) *)

*) For the case $P_0=0$ and c a positive integer, G.Margreiter has derived special formulas, cf. /53/ .

sin and cos: $q = \sin p$ $\qquad Q_\mu = \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)R_\rho P_{\mu-\rho}$, $Q_0 = \sin P_0$

$r = \cos p$ $\qquad R_\mu = -\frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)Q_\rho P_{\mu-\rho}$, $R_0 = \cos P_0$

Proofs can be found in Moore /36/ and Wanner /51/.

When all operations that give the functions $f_i(x,y_1,\ldots,y_n)$ from x and $y_1,\ldots,y_n$ are replaced by the corresponding recursion formulas, these will give the values $F_{i\mu}$ from X_μ and $Y_{1,\mu},\ldots,Y_{n\mu}$ if all derivatives of lower order are still present. Because of

(2.4) $\qquad Dy_i = f_i$, $D^{\mu+1}y_i = D^\mu f_i$

these quantities are equal to $F_{i\mu} = (\mu+1)Y_{i,\mu+1}$. Hence,

(2.5) $\qquad Y_{i,\mu+1} = \frac{1}{\mu+1}F_{i\mu}$, $\qquad (\mu=0,1,2,\ldots)$ $\qquad (i=1,2,\ldots)$.

The procedure can now be repeated with the quantities $Y_{i,\mu+1}$. It will lead to a recusive computation of the $Y_{i\mu}$. Recursion begins with the values $Y_{i0}=y_i$ (initial values) using the formulas

(2.6) $\quad X_0=x,\ X_1=1,\ X_2=X_3=\ldots=0$

and, for a constant c,

(2.7) $\quad C_0=c,\ C_1=C_2=\ldots=0$

Then, it proceeds according to the pattern

$X_0,Y_{10},\ldots,Y_{n0} \rightarrow \ldots P_0,Q_0 \rightarrow R_0 \ldots \rightarrow F_{10},\ldots,F_{n0}$

$Y_{11},\ldots,Y_{n1} \rightarrow \ldots P_1,Q_1\ R_1 \ldots \rightarrow F_{11},\ldots,F_{n1}$

$Y_{12},\ldots,Y_{n2}\quad \ldots P_2,Q_2\ R_2 \ldots\quad F_{12},\ldots,F_{n2}$

$Y_{13},\ldots,Y_{n3}$ etc.

Some of the authors that have worked with one of these (or similar) recursion formulas are Steffensen /46/, Miller-Hurst /35/, E.Rabe /39/, W.Gautschi /16/, E.Fehlberg /14/, I.Jennig / /, Deprit-Zahar / 9/, Leavitt /32/, Richtmyer /41/.

II.3. Estimation of Error

Estimating the error of a series that has been cut off (e.g., after the m-th term) is thus indispensable for a sensible choice and control of the step size. One possibility is to bound the error by means of majorant series (e.g., W. Groebner /21/, /22/).
However, with Duffing's differential equation as an example, G. Maeß /34/ has shown that using Lagrange's remainder of Taylor's series gives an error limit which is by 3-4 powers of ten more accurate than in the case of the majorant technique.
If all occuring derivatives exist and are continuous, then we have, according to Lagrange,

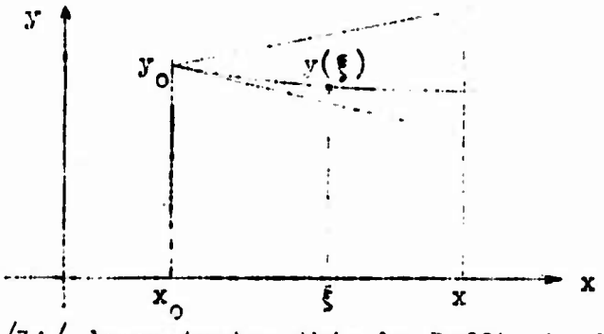
$$y_i(x) = \sum_{\mu=0}^{m} \frac{(x-x_0)^\mu}{\mu!} \left[D^\mu y_i \right]_{x_0,y_0} + R_{im}$$

with $$R_{im} = \frac{(x-x_0)^{m+1}}{(m+1)!} \left[\frac{d^{m+1}}{dx^{m+1}} y_i(x) \right]_{x=\xi}, \quad x_0 \leq \xi \leq x$$

where, owing to (1.2')

$$R_{im} = \frac{(x-x_0)^{m+1}}{(m+1)!} \left[D^{m+1} y_i \right]_{\xi, y(\xi)}, \quad x_0 \leq \xi \leq x \tag{3.1}$$

For a precise estimation of the error one has to know a domain B which is known to contain the solution $y(\xi)$. The functions $D^{\mu+1} y_i$ can then be estimated in this domain (Fig. 1)



(Fig. 1)

Maeß /34/ demonstrates this by Duffing's differential equation.
R.E. Moore /36/ solves this problem generally and automatically by means of interval arithmetics.

knowing an approximate error is sufficient for a sensible control of the step size. Here, one may put up with, say, the value of $D^{n+1}y_i$ at the point x_o, y_o (this would be the first term neglected), or rather: one chooses the larger one of the values at the points x_o, y_o and $x_o+h, \hat{y}(x_o+h)$ (starting point for the subsequent step). Both numbers are easy to compute: it is sufficient to run the iteration for calculating the Taylor coefficients for this and the next step through another loop.

We shall obtain the formula (3.1) for the remainder as a special case in the next chapter.

II.4. Transfer Matrices

Let $y_i(x)$ be solutions of the differential equation (I.1.1) for the initial values y_{ko}. The matrix

$$H(x) = \left(H_{ik}(x)\right) = \left(\frac{\partial y_i(x)}{\partial y_{ko}}\right) \tag{4.1}$$

which consists of the <u>derivatives</u> of the solution $y_i(x)$ with respect to the k-th initial value y_{ko}, is then called the <u>transfer matrix</u> pertaining to $y(x)$.

In other words: The transfer matrix describes, in first approximation, the variation of the solutions y_i at the point x if the initial values y_{ko} are changed. When we change the initial values $y_{1o},\dots,y_{no}$ by $\varepsilon_{1o},\dots,\varepsilon_{no}$, the solutions y_i at the point x will in first approximation change for

$$\varepsilon_i(x) = \frac{\partial y_i(x)}{\partial y_{1o}}\,\varepsilon_{1o} + \dots + \frac{\partial y_i(x)}{\partial y_{no}}\,\varepsilon_{no} \quad .$$

Thus,

$$\begin{pmatrix} \varepsilon_1(x) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial y_{1o}} & \dots & \frac{\partial y_1(x)}{\partial y_{no}} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial y_n(x)}{\partial y_{1o}} & \dots & \frac{\partial y_n(x)}{\partial y_{no}} \end{pmatrix} \begin{pmatrix} \varepsilon_{1o} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{no} \end{pmatrix} \tag{4.2}$$

or, in vectorial form,

(4.2') $\varepsilon(x) = H(x)\varepsilon_o$.

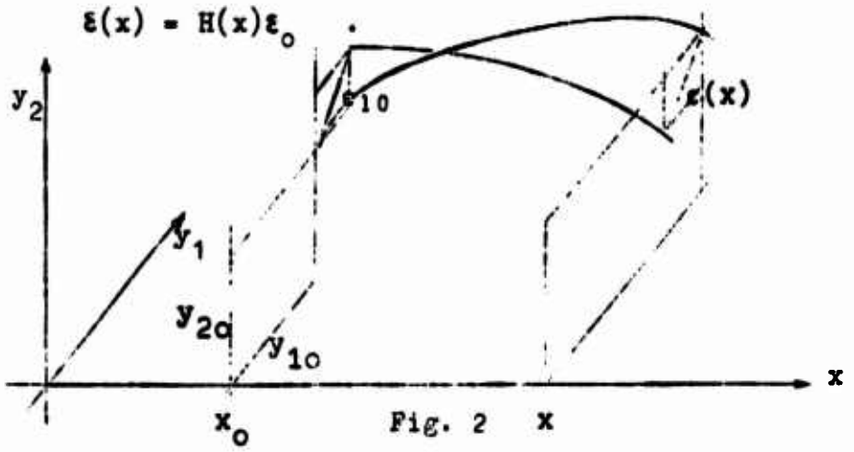


Fig. 2

For example, errors committed somewhere in the numerical integration can be mapped forward or backward to any fixed point by means of the transfer matrices.

Hence, these also describe how an error committed at a certain place influences the final result. We shall consult the transfer matrices also for an "optimum" step size control which takes stability and the total final error into account (Chapter VI).

The transfer matrices are also useful in boundary value problems in which some of the initial values are missing and have been replaced by conditions at other parametric points. Here, the missing initial values must first be guessed and then be improved by means of the transfer matrices, after the relevant solutions have been found (Wanner /51/).

II.5. Calculation of the Transfer Matrices

To calculate the transfer matrix preliminarily for a small domain, we differentiate the solution series (1.5) term by term with respect to the initial value y_{ko}:

$$(5.1) \qquad H_{ik}(x) = \frac{\partial y_i(x)}{\partial y_{ko}} = \sum_{\mu=0}^{\infty} \frac{(x-x_o)^\mu}{\mu!} \left[\frac{\partial}{\partial y_k} D^\mu y_i \right]_{x_o, y_o} .$$

In the next section, we shall find recursion formulas for the calculation of the expressions $\frac{\partial}{\partial y_k} D^\mu y_i$.

A remainder formula for the error after the m-th term, which is analogous to (3.1), is

$$(5.2)\qquad S_{ikm}(x)=\frac{(x-x_o)^{m+1}}{(m+1)!}\left\{\sum_{j=1}^{n}\left[\frac{\partial}{\partial y_j}D^{m+1}y_i\right]_{\xi,y(\xi)}\frac{\partial y_j(\xi)}{\partial y_{ko}}\right\}$$

$$x_o\leq\xi\leq x\quad .$$

In the case of a step-by-step integration of the differential equations with the intervals $x_o<x_1<\ldots<x_N$, Eq. (5.1) gives only the local transfer matrices

$$C(x_j,x_{j-1})=\left(\frac{\partial y_i(x_j)}{\partial y_k(x_{j-1})}\right)\quad .$$

Owing to the chain rule (for functions of several variables), these matrices can be multiplied with each other to give the total transfer matrix

$$(5.3)\qquad H(x_N)=C(x_N,x_{N-1})\ldots C(x_2,x_1)\,C(x_1,x_o)$$

Notice that

$$(5.4)\qquad H(x_o)=C(x,x)=E\qquad\text{(Identity matrix)}$$

and

$$(5.5)\qquad C(x,x')=C(x',x)^{-1}\quad .$$

For linear systems of differential equations, the columns of the transfer matrix coincide with the fundamental solutions of the corresponding homogeneous system (with the initial values $0,..,1,..,0$), and the relation (4.2) not only holds in first approximation but is valid exactly.

Another possible way of calculating the transfer matrix is to integrate the system

$$\frac{dH_{ik}(x)}{dx}=\sum_{j=1}^{n}\left[\frac{\partial f_i}{\partial y_j}\right]_{x,y(x)}H_{jk}(x)$$

for every $k=1,\ldots,n$ with the initial values

$$H_{ik}(x_o)=\delta_{ik}$$

together with Eq. (I.1.1). This formula is usually given.

II.6. Recursion Formulas for Expressions with an Additional Operator

Here, we replace the operator $\frac{\partial}{\partial y_k}$ of (5.1) generally by $\bar{D}$ because we shall need the following formulas for other purposes too. Suppose $\bar{D}$ is another linear differential operator. In addition to (2.3) we adopt the symbols

$$(6.1) \qquad \bar{P}_\mu = \frac{\bar{D}D^\mu p}{\mu!} \quad , \quad \bar{Q}_\mu = \frac{\bar{D}D^\mu q}{\mu!} \quad , \quad \bar{R}_\mu = \frac{\bar{D}D^\mu r}{\mu!}$$

for certain operands p,q,r. Again, these quantities are functions of $x,y_1,\dots,y_n$, and the corresponding symbols $\bar{X}_\mu,\bar{Y}_{1\mu},\dots,\bar{Y}_{n\mu},\bar{F}_{1\mu},\dots,\bar{F}_{n\mu}$, $\bar{C}_\mu$ are again valid for the functions $x,y_1,\dots,y_n,f_1,\dots,f_n,c$.
Also for these quantities we obtain recursion formulas by simply applying the operator $\bar{D}$ to the formulas of page

Table

Sum: r=p+q	$\bar{R}_\mu=\bar{P}_\mu+\bar{Q}_\mu$
Difference: r=p-q	$\bar{R}_\mu=\bar{P}_\mu-\bar{Q}_\mu$
Product r=p.q	$\bar{R}_\mu=\sum_{\rho=0}^{\mu}\bar{P}_\rho Q_{\mu-\rho}+P_\rho\bar{Q}_{\mu-\rho}$
Quotient: r=p/q	$\bar{R}_\mu=\{\bar{P}_\mu-\sum_{\rho=0}^{\mu-1}(\bar{R}_\rho Q_{\mu-\rho}+R_\rho\bar{Q}_{\mu-\rho})-R_\mu\bar{Q}_0\}/Q_0$
exp: r=exp p	$\bar{R}_\mu=\frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)\{\bar{R}_\rho P_{\mu-\rho}+R_\rho\bar{P}_{\mu-\rho}\}$, $\bar{R}_0=R_0\bar{P}_0$
log r=log p	$\bar{R}_\mu=\{\bar{P}_\mu-\frac{1}{\mu}\sum_{\rho=1}^{\mu-1}(\mu-\rho)(\bar{P}_\rho R_{\mu-\rho}+P_\rho\bar{R}_{\mu-\rho})-\bar{P}_0R_\mu\}/P_0$, $\bar{R}_0=\bar{P}_0/P_0$
Root: r= p	$\bar{R}_\mu=\frac{1}{2R_0}\{\bar{P}_\mu-2\sum_{\rho=0}^{\mu-1}\bar{R}_\rho R_{\mu-\rho}\}$ $\mu=0,1,\dots$
Constant power: $r=p^c$	$\bar{R}_\mu=\{\frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(c\mu-(c+1)\rho)(\bar{R}_\rho P_{\mu-\rho}+R_\rho\bar{P}_{\mu-\rho})-R_\mu\bar{P}_0\}/P_0$, $\bar{R}_0=cR_0\bar{P}_0/P_0$ $P_0\neq0$

sin, cos q=sin p

$$\bar{Q}_\mu = \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)\{\bar{R}_\rho P_{\mu-\rho}+R_\rho \bar{P}_{\mu-\rho}\}\,, \quad \bar{Q}_0=R_0\bar{P}_0$$

r=cos p

$$\bar{R}_\mu = \frac{-1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)\{\bar{Q}_\rho P_{\mu-\rho}+Q_\rho \bar{P}_{\mu-\rho}\}\,, \quad \bar{R}_0=-Q_0\bar{P}_0$$

See Wanner /51/, p. 27.

First of all, all expressions R_μ must exist should $\bar{R}_\mu$ be calculated. Applying $\bar{D}$ to (2.5) we obtain

$$\bar{Y}_{i,\mu+1} = \frac{1}{\mu+1}\bar{F}_{i\mu} \tag{6.2}$$

which enables us to employ recursion. As we can see, the above formulas are independent of the particular choice of the operator $\bar{D}$. Setting, e.g., $\bar{D} = \frac{\partial}{\partial y_k}$ we find the expressions

$$\bar{Y}_{i\mu} = \frac{1}{\mu!}\frac{\partial}{\partial y_k} D^\mu y_i$$

which are needed in (5.1). In this case, recursion starts with the initial values

$$\bar{Y}_{io} = \frac{\partial}{\partial y_k} y_i = \begin{cases} 1 & i=k \\ 0 & i\neq k \end{cases}$$

For the independent variable x we have $\bar{X}_0 = \bar{X}_1 = \ldots = 0$

and for a constant c $\quad \bar{C}_0 = \bar{C}_1 = \ldots = 0$.

Subroutines, which calculate these formulas are given in Knapp-Wanner /30/ or Wanner /51/.

II.7. Recursion Formulas for Other Operations

The class of operations that are allowed for the formation of the functions $f_i(x,y)$ will be considerably expanded in this section. We shall see that every function satisfying a differential equation that can already be processed is allowed here.

First, we show by way of a few examples how recursion formulas can be

got for many functions by introducing auxiliary expressions:

r=arctan p :

Here, $$Dr = \frac{Dp}{1+p^2} \quad .$$

We set $1+p^2=q$, whence $qDr=Dp$.

We find

$$Q_\mu = \sum_{\rho=0}^{\mu} P_\rho P_{\mu-\rho} \quad , \quad Q_0 = 1+P_0^2$$

(7.1)

$$R_\mu = \Big\{P_\mu - \frac{1}{\mu}\sum_{\rho=1}^{\mu-1}(\mu-\rho)Q_\rho R_{\mu-\rho}\Big\}/Q_0 \quad , \quad R_0 = \arctan P_0 \quad .$$

r=tan p : When sin p and cos p occur simultaneously, the best way is to write $r = \frac{\sin p}{\cos p}$ and to use the formulas of page .

When sin p or cos p does not occur, it is preferable to use the formulas

$$Q_{\mu-1} = \sum_{\rho=0}^{\mu-1} R_\rho R_{\mu-\rho-1} \quad , \quad Q_0 = 1+R_0^2$$

(7.2)

$$R_\mu = \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)Q_\rho P_{\mu-\rho} \quad , \quad R_0 = \tan P_0$$

which have been obtained by reversing the formulas (7.1).

r=arcsin p :

We set $q=\sqrt{(1-p^2)}$, whence $qDr=Dp$. Owing to $q^2=1-p^2$ we obtain

$$Q_\mu = \frac{-1}{2Q_0}\Big[\sum_{\rho=0}^{\mu} P_\rho P_{\mu-\rho} + \sum_{\rho=1}^{\mu-1} Q_\rho Q_{\mu-\rho}\Big] \quad , \quad Q_0 = \sqrt{(1-P_0^2)}$$

(7.3)

$$R_\mu = \Big\{P_\mu - \frac{1}{\mu}\sum_{\rho=1}^{\mu-1}(\mu-\rho)Q_\rho R_{\mu-\rho}\Big\}/Q_0 \quad , \quad R_0 = \arcsin P_0 \quad .$$

For r=arccos p , all formulas remain the same, except for $R_0=\arccos P_0$

For the corresponding hyperbolic functions, only a few signs have to be changed in the formulas on page 16. Of course, also for all these formulas there are also the corresponding recursions with the additional operator $\bar{D}$.

Consider the general case that $u_1(x),\ldots,u_m(x)$ are solutions of the differential equations

$$u_i'(x) = g_i(x,u(x)) \ .$$

The only assumption we make is that the functions g_i are made up only of the operations dealt with so far. Then we can give recursion formulas also for those functions. This step can be repeated over and over and leads to a successive extension of the recursion formulas to more and more functions of analysis.

Let

$$r_i = u_i(p) \qquad (i=1,\ldots,m)$$

Since the functions g_i are made up of operations whose recursions are known, we can calculate the expressions

$$G^*_{i,\mu-1} = \frac{D^{\mu-1}g_i(p,u(p))}{(\mu-1)!}$$

for the operand $p(x,y)$ from the coefficients up to $P_{\mu-1}$, $R_{i,\mu-1}$.

Because of $Dr_i=u_i'(p)Dp = g_i(p,u(p))Dp$ we have

$$(7.4) \qquad R_{i\mu} = \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)G^*_{i\rho}P_{\mu-\rho} \ , \quad R_{io}=u_i(P_o) \ ,$$

the sought recursion formula.

Finally, we consider the important equation

$$(7.5) \qquad a_3(x)u'' + a_2(x)u' + a_1(x)u = 0$$

which with the usual substitutions $u=u_1$, $u'=u_2$ becomes

$$u_1' = u_2$$

$$u_2' = \frac{a_1u_1+a_2u_2}{a_3}$$

Let $p(x,y)$ be an arbitrary operand and let

$$r_1=u_1(p)=u(p) \ , \quad r_2=u_2(p)=u'(p) \ :$$

We put $a_k(p)=a_k^*(x,y)$ and assume that the coefficients

$$A^*_{k\rho} = \frac{D^\rho a_k(p)}{\rho!}$$

can be calculated by means of the existing recursion formulas from the values $P_0,\dots,P_\rho$.

Moreover, we use the notation

$$a_1(p)u_1(p) + a_2(p)u_2(p) = a^*_1 r_1 + a^*_2 r_2 = q \quad , \quad \frac{q}{a^*_3} = s \; ;$$

then we find the recursion formulas

$$\left.\begin{aligned}
Q_{\mu-1} &= \sum_{\rho=0}^{\mu-1}\left[A^*_{1\rho}R_{1,\mu-\rho-1}+A^*_{2\rho}R_{2,\mu-\rho-1}\right] \\
S_{\mu-1} &= \left[Q_{\mu-1} - \sum_{\rho=0}^{\mu-2} S_\rho A^*_{3,\mu-\rho-1}\right]/A^*_{3,0} \\
R_{1\mu} &= \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)R_{2\rho}P_{\mu-\rho} \\
R_{2\mu} &= \frac{1}{\mu}\sum_{\rho=0}^{\mu-1}(\mu-\rho)S_\rho P_{\mu-\rho}
\end{aligned}\right\} \qquad (\mu=1,2,\dots) \tag{7.6}$$

$$R_{1,0} = u(P_0) \quad , \quad R_{2,0} = u'(P_0) \quad .$$

These recursions are valid for all functions that satisfy a differential equation of the form (7.5), that is, for example, all kinds of Bessel functions, Mathieu functions, Weber functions, Chebyshev-, Legendre-, Hermite-, Laguerre-, or Jacobi polynomials, etc.
For Bessel functions of the first kind, e.g., we have

$$a^*_3 = p^2 \quad , \quad a^*_2 = p \quad , \quad a^*_1 = p^2-n^2$$

and

$$A^*_{3\mu} = \sum_{\rho=0}^{\mu} P_\rho P_{\mu-\rho} \quad , \quad A^*_{2\mu} = P_\mu \quad , \quad A^*_{1\mu} = A^*_{3\mu} - n^2\delta_{\mu 0} \; .$$

Finally, we should like to mention that for orthogonal polynomials, in particular for higher n (low n are uninteresting), the above formulas require much less work than using the "generating functions" as suggested by Leavitt /32/.

Chapter III

LIE-Series

by G. Wanner

This chapter discusses the numerical evaluation of W. Groebner's Lie series perturbation formula, on which an efficient numerical method with satisfactory error estimation is based.

III.1. Groebner's Perturbation Formula

Groebner's perturbation formula states how one has to correct an arbitrary given approximate solution $\hat{y}(x)$ $(\hat{y}_1(x),\ldots,\hat{y}_n(x))$ in order to find the solution $y(x)$. This formula is a generalization of Taylor's series (II.1.5) which can be obtained from it when the operators are chosen in a special way.
A system of differential equations must be known for the approximate solution $\hat{y}(x)$:

$$\hat{y}' = \hat{f}(x,\hat{y}) \tag{1.1}$$

and the approximate solution must assume the same initial values

$$\hat{y}(x_0) = y_0 \quad . \tag{1.2}$$

We introduce the operator

$$D_2 = D - D_1 = (f_1(x,y) - \hat{f}_1(x,y))\frac{\partial}{\partial y_1} + \ldots + (f_n(x,y) - \hat{f}_n(x,y))\frac{\partial}{\partial y_n} \tag{1.3}$$

which accounts for the difference between the two differential equations. Hence,

$$D = D_1 + D_2 \tag{1.4}$$

As we shall see in the next chapter, here we have the formula

$$y(x) = \hat{y}(x) + \sum_{\alpha=0}^{\infty} \int_{x_0}^{x} \frac{(x-\xi)^\alpha}{\alpha!} \left[D_2 D^\alpha y\right]_{\xi,\hat{y}(\xi)} \, d\xi \tag{1.5}$$

(W. Groebner) for the sought solution $y(x)$.

III.2. Knapp's Remainder Formula

According to Knapp /26/, the remainder of the series (1.5) after, say, s terms is

$$y(x) = \bar{y}(x) + R_s(x) \tag{2.1}$$

with

$$(2.1') \qquad \bar{y}(x) = \hat{y}(x) + \sum_{\alpha=0}^{s} \int_{x_0}^{x} \frac{(x-\xi)^\alpha}{\alpha!} \left[D_2 D^\alpha y\right]_{\xi,\hat{y}(\xi)} d\xi$$

$$(2.1'') \qquad R_s(x) = \int_{x_0}^{x} \frac{(x-\xi)^s}{s!} \left\{\left[D^{s+1}y\right]_{\xi,y(\xi)} - \left[D^{s+1}y\right]_{\xi,\hat{y}(\xi)}\right\} d\xi$$

We shall prove this formula in the next chapter. Knapp / 26 / has derived these formulas assuming that $f_i, \hat{f}_i \in C^s$.

Another formula for the error can be obtained by increasing s in (2.1'') by 1 and adding the last term $\alpha=s+1$ of (2.1'):

$$(2.2) \qquad R_s(x) = \int_{x_0}^{x} \frac{(x-\xi)^{s+1}}{(s+1)!} \left\{\left[DD^{s+1}y\right]_{\xi,y(\xi)} - \left[D_1 D^{s+1}y\right]_{\xi,\hat{y}(\xi)}\right\} d\xi$$

Finally, a mean value theorem of integral calculus gives

$$(2.3) \qquad R_s(x) = \frac{(x-x_0)^{s+1}}{(s+1)!} \left\{\left[D^{s+1}y\right]_{\xi_1,y(\xi_1)} - \left[D^{s+1}y\right]_{\xi_1,\hat{y}(\xi_1)}\right\}$$

$$x_0 \leq \xi_1 \leq x$$

and

$$(2.4) \qquad R_s(x) = \frac{(x-x_0)^{s+2}}{(s+2)!} \left\{\left[DD^{s+1}y\right]_{\xi_2,y(\xi_2)} - \left[D_1 D^{s+1}y\right]_{\xi_2,\hat{y}(\xi_2)}\right.$$

$$x_0 \leq \xi_2 \leq x$$

III.3. Special Case: Power Series

The power series expansion of the solution y(x) is a special case of the Lie series (2.1), if the original differential equation is autonomous, i. e., if the function f(x,y) do not depend on x and we have $D = \frac{\partial}{\partial x} + f(y)\frac{\partial}{\partial y}$.

To show this we put $D_1 = \frac{\partial}{\partial x}$, $D_2 = f(y)\frac{\partial}{\partial y}$, $\hat{y}(x)=y_0$, thus

$\left[D_2 D^\alpha y\right]_{\xi,\hat{y}(\xi)} = \left[D^{\alpha+1}y\right]_{x_0,y_0}$ since here also $D^\alpha y$ do not depend on

Now the integrations are readily carried out giving with $\alpha+1=\beta$ and the remainder (2.4)

$$y(x) = \sum_{\beta=0}^{s+1} \frac{(x-x_0)^\beta}{\beta!}\left[D^\beta y\right]_{x_0,y_0} + \frac{(x-x_0)^{s+2}}{(s+2)!}\left[D^{s+2}y\right]_{\xi_2,y(\xi_2)} \quad , \quad x_0 \leq \xi_2 \leq x$$

the formula (II.3.1) which has been used before.

III.4. Choice of Approximate Solutions

The approximate solutions $\hat{y}_1(x),\dots,\hat{y}_n(x)$ can be chosen freely. They only have to satisfy the initial conditions (1.2), and a system of differential equations must be known for them. The better the choice of the approximate solutions, the more efficient is the method.
It is expedient to use the first terms of the power series expansion (II.1.5) for an approximation to start with:

$$(4.1) \qquad \hat{y}_i(x) = \sum_{\mu=0}^{m} \frac{(x-x_0)^\mu}{\mu!}\left[D^\mu y_i\right]_{x_0,y_0} = \sum_{\mu=0}^{m} (x-x_0)^\mu \left[Y_{i\mu}\right]_{x_0,y_0}$$

(cf. (II.2.3)); of course, m may also depend on i. A corresponding system of differential equations can be found by simply differentiating Eq. (4.1) (the quantities $\left[Y_{i\mu}\right]_{x_0,y_0}$ are constants)

$$(4.1') \qquad \hat{f}_i(x,y)=\hat{y}_i'(x)=\sum_{\mu=1}^{m} (x-x_0)^{\mu-1}\left[\mu Y_{i\mu}\right]_{x_0,y_0} = \sum_{\mu=0}^{m-1} (x-x_0)^\mu \left[F_{i\mu}\right]_{x_0,y_0}$$

(cf. (II.2.5)), where the functions $\hat{f}_i$ depend on x only. The formulas (4.1) and (4.1') are used in the general program GROEBNER, reproduced in Knapp-Wanner /30/ or Wanner /51/.

One may also retain parts of the original system, e.g., in equations of the kind

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= y_3 \\ &\dots \\ y_n' &= f(x,y_1,\dots,y_n) \end{aligned}$$

and replace only the last equation by a polynomial in x. In this case, however, the degrees m in Eq. (4.1) must decrease by unity each as i increases. This reduces the operator D_2 to a simpler form since then it consists merely of a single term.

When the functions $D^{m+1}y_i$ are bounded in a region B of (x,y)-space, then we have from (II.3.1)

$$|y_i(x)-\hat{y}_i(x)| \leq C \frac{|x-x_o|^{m+1}}{(m+1)!} \tag{4.2}$$

or

$$y_i(x)-\hat{y}_i(x) = O((x-x_o)^{m+1}) \quad . \tag{4.2'}$$

In this case we say that $\hat{y}_i(x)$ is <u>of the order m</u> (or of the <u>error order m+1</u>).

Of course there are examples for which a choice other than (4.1) is more convenient, e.g., the equation

$$y' = \sqrt{x} + \sqrt{y} \quad , \quad y(0) = 0 \quad .$$

Here, the first Taylor term vanishes, the second is infinite. However, choosing the approximate solution

$$\hat{y}' = \sqrt{x} \quad , \quad \hat{y} = \frac{2}{3} x^{3/2}$$

we obtain from (2.1') with s=0

$$\bar{y} = \frac{2}{3} x^{3/2} + \sqrt{\frac{2}{3}} \frac{4}{7} x^{7/4} \quad .$$

Compared with other methods, this is a very good approximation /42/. For small x values its accuracy is sufficient and the singular point x=0 can be avoided. More terms of Eq. (2.1') are not allowed, because only $f \in C^0$, whereas $f \notin C^1$.

III.5. Order of the Method

<u>Definition:</u>

A method is of the <u>order</u> p, if the solutions $\hat{y}(x)$ obtained through it are of the order p, i.e., if for <u>every</u> solution y(x), whose Taylor series exists far enough,

$$y_i(x) - \hat{y}_i(x) = O((x-x_0)^{p+1}) \quad .$$

The Taylor series of the two solutions will then agree up to at least the p-th term.

We shall prove now that the method defined by Eq. (2.1') is of the order m+s+1, if the starting solution $\hat{y}(x)$ is of the order m:

Theorem: If the functions $D^s f_i$ satisfy a Lipschitz condition

$$(5.1) \qquad \left|\left[D^s f_i\right]_{x,y^*} - \left[D^s f_i\right]_{x,y^{**}}\right| \le K_s \sum_{k=1}^{n} |y_k^* - y_k^{**}|$$

in a region B, and if the starting solution is of the order m

$$(5.2) \qquad |y_i(x) - \hat{y}_i(x)| \le M \frac{|x-x_0|^{m+1}}{(m+1)!} \quad ,$$

then the relation

$$(5.3) \qquad |y_i(x) - \bar{y}_i(x)| \le M \frac{K_s n |x-x_0|^{m+s+1}}{(m+s+1)!}$$

holds true for the solution $\bar{y}(x)$ that follows from (2.1').

Proof: $D^{s+1} y_i = D^s f_i$, hence substituting (5.1) and (5.2) in (2.1'') we get

$$|y_i(x) - \bar{y}_i(x)| = |R_{is}(x)| =$$

$$\left|\int_{x_0}^{x} \frac{(x-\xi)^s}{s!} \left\{\left[D^{s+1} y_i\right]_{\xi, y(\xi)} - \left[D^{s+1} y_i\right]_{\xi, \hat{y}(\xi)}\right\} d\xi \right| \le$$

$$\le \left|\int_{x_0}^{x} \frac{|x-\xi|^s}{s!} K_s \sum_{k=1}^{n} |y_k(\xi) - \hat{y}_k(\xi)| \, d\xi \right| \le$$

$$\le \left|\int_{x_0}^{x} \frac{|x-\xi|^s}{s!} K_s n M \frac{|\xi - x_0|^{m+1}}{(m+1)!} \, d\xi \right| \quad .$$

Now the statement follows by means of the well-known integral formula

$$(5.4) \qquad \frac{(x-x_0)^{\mu}}{\mu!} = \int_{x_0}^{x} \frac{(x-\xi)^{\alpha}}{\alpha!} \frac{(\xi - x_0)^{\mu-\alpha-1}}{(\mu-\alpha-1)!} \, d\xi \quad , \qquad (0 \le \alpha \le \mu - 1) \quad .$$

This theorem is a special case of a general theorem stated in the

next chapter.

Thus, the order of the method increases by 1 with each additional integral. It may also increase by more than 1, as for example in the following case:

$$y' = x^{100}+y^{100} \quad , \quad y(0) = 0 \quad ,$$

$$D = \frac{\partial}{\partial x} + (x^{100}+y^{100})\frac{\partial}{\partial y} \quad , \quad D_1 = \frac{\partial}{\partial x} + x^{100}\frac{\partial}{\partial y} \quad , \quad D_2 = y^{100}\frac{\partial}{\partial y}$$

$$\hat{y}(x) = \frac{x^{101}}{101}$$

where Eq. (2.1') with s=1 gives

$$y(x) = \hat{y}(x) + \int_0^x \frac{\xi^{10100}}{(101)^{100}}\,d\xi + \int_0^x (x-\xi)\,100\,\frac{\xi^{20099}}{(101)^{199}}\,d\xi + \dots$$

$$= \frac{x^{101}}{101} + \frac{x^{10101}}{10101(101)^{100}} + \frac{x^{20101}}{20101.201.(101)^{199}} + \dots \ ;$$

this contains already more than 30 000 Taylor terms.

III.6. Numerical Evaluation, Quadrature Formulas

To evaluate Eq. (2.1') numerically we must calculate the occuring integrals in a proper way. The following lemma is quite useful for this purpose.

Lemma: If the starting solution $\hat{y}(x)$ is of the order m

$$y_i(x) - \hat{y}_i(x) = O((x-x_o)^{m+1})$$

and if the $f_i(x,y)$ satisfy a Lipschitz condition, then

$$(6.1) \qquad f_k(x,\hat{y}(x)) - \hat{f}_k(x,\hat{y}(x)) = O((x-x_o)^m) \quad .$$

Proof: According to the first assumption we have *)

$$y_i'(x) - \hat{y}_i'(x) = O((x-x_o)^m)$$

---

*) because $y_i(x)-y_i(x) \in C^{m+1}$

and owing to the Lipschitz condition we have

$$f_i(x,\hat{y}(x)) - f_i(x,y(x)) = O((x-x_o)^{m+1}) \quad .$$

Hence,

$$f_i(x,\hat{y}(x)) - \hat{f}_i(x,\hat{y}(x)) =$$

$$= f_i(x,\hat{y}(x)) - f_i(x,y(x)) + f_i(x,y(x)) - \hat{f}_i(x,\hat{y}(x)) =$$

$$= O((x-x_o)^{m+1}) + y_i'(x) - \hat{y}_i'(x) =$$

$$= O((x-x_o)^m) \quad .$$

Now we have to calculate the following integrals

$$(6.2) \qquad \int_{x_o}^{x} \frac{(x-\xi)^\alpha}{\alpha!} \left[D_2 D^\alpha y_i\right]_{\xi,\hat{y}(\xi)} d\xi = \int_{x_o}^{x} g(\xi)d\xi \quad .$$

We choose the step size h and set, as usual according to Gauss,

$$(6.3) \qquad \int_{x_o}^{x_o+h} g(\xi)d\xi = h\sum_{j=1}^{k} c_j g(x_o+a_j h)$$

where the a_j $(0 \le a_j \le 1)$ determine the given basic points at which the values of $g(\xi)$, which are then summed up with the weights, must be calculated. The rest of this section will now be dedicated to determining the coefficients a_j and c_j as expediently as possible.
First, we find from the lemma that the function $g(\xi)$ contains the factor $(\xi-x_o)^m$, for we have (cf. (1.3))

$$\left[D_2 D^\alpha y_i\right]_{\xi,\hat{y}(\xi)} = \sum_{j=1}^{n} \left\{f_j(\xi,\hat{y}(\xi)) - \hat{f}_j(\xi,\hat{y}(\xi))\right\} \left[\frac{\partial}{\partial y_j} D^\alpha y_i\right]_{\xi,\hat{y}(\xi)} ,$$

hence, owing to (6.1), $(\xi-x_o)^m$ is a factor occurring in the braces.

Thus,

$$g(\xi) = (\xi-x_o)^m G(\xi)$$

and (6.3) attains the form

$$(6.4) \qquad \int_{x_o}^{x_o+h} (\xi-x_o)^m G(\xi)d\xi = h^{m+1} \sum_{j=1}^{k} c_j a_j^m G(x_o+a_j h) \quad .$$

The transformation $\xi-x_o=ht$ gives

$$(6.5) \qquad \int_0^1 t^m G^*(t)dt = \sum_{j=1}^{k} C_j G^*(a_j)$$

with

$$(6.6) \qquad G^*(t) = G(x_0+ht)$$

and

$$(6.7) \qquad C_j = c_j a_j^m \quad .$$

Equation (6.5) shows how the coefficients C_j and a_j must be determined so that an order as high as possible will be attained: The a_j must be the zeros of the k-th one of the polynomials which in the interval (0,1) are orthogonal with the weight function t^m, the C_j are the corresponding weights (e.g., Natanson /38/, p. 436). These coefficients are tabulated with 8D, e.g., in Krylov-Lugin-Janovich /31/. Stroud-Secrest /49/ give a FORTRAN program for this (however, for the interval (-1,+1)). The coefficients c_j can then be found by means of (6.7). They can be calculated explicitly for k=1,2:

$$k=1: \quad a_1 = \frac{m+1}{m+2} \quad , \qquad c_1 = \frac{1}{(m+1)a_1^m}$$

$$k=2: \quad a_{1,2} = \frac{m+2 \pm \sqrt{2(m+2)/(m+3)}}{m+4}$$

$$c_1 = \left(\frac{1}{m+2} - a_2 \frac{1}{m+1}\right) / \left(a_1^m(a_1-a_2)\right)$$

$$c_2 = \left(\frac{1}{m+2} - a_1 \frac{1}{m+1}\right) / \left(a_2^m(a_2-a_1)\right) \quad .$$

III.7. Some Values of the Table of Coefficients

Here are the coefficients a_j, c_j of the quadrature formula (6.3) for a few values of m and for k=1(1)4 with an accuracy of about 25 places.

		A	C
I=0			
	K=1	5.0000000000000000000000 -001	1.00000000000000000000000-000
	K=2	7.8867513459481288225457 5-001	5.00000000000000000000000-001
		2.1132486540518711774542 6-001	5.00000000000000000000000-001
	K=3	8.8729833462074168851792 7-001	2.77777777777777777777778-001
		5.0000000000000000000000-001	4.44444444444444444444444-001
		1.1270166537925831148207 3-001	2.77777777777777777777778-001
	K=4	9.3056815579702628761197 3-001	1.73927422568726928686532-001
		6.6999052179242813240133 3-001	3.26072577431273071313468-001
		3.3000947820757186759866 7-001	3.26072577431273071313468-001
		6.9431844202973712388026 7-002	1.73927422568726928686532-001
M=5		A	C
	K=1	6.5714285714285714285714 3-001	3.60232330820301783264746-001
	K=2	9.2476396172581052836120 1-001	1.89745147231564037755387-001
		6.3079159582974496719435 5-001	3.10387958274777127577698-001
	K=3	9.5210976664102754544813 -001	1.21065093877915067765459-001
		7.6162396969946054247251 0-001	2.52061620666265055584747-001
		4.6798323545461852116405 -001	3.25082120967065191102105-001
	K=4	9.6653864646511777190977 2-001	8.49165021362317656076087-002
		8.3107900386011409851134 0-001	1.31744296965435662580849-001
		6.1466938985537839249657 6-001	2.44220383106226852185624-001
		3.5609372905015896785154 3-001	2.65012632244661778769339-001

		A	C
M=10			
	K=1	9.16666666666666666666667-001	2.17016584324794239848481 001
	K=2	9.54195174355251064338066-001	1.1719924042989081967535[illegible]-001
		7.60090539930463221376219-001	2.73073421491466730325113[illegible]-001
	K=3	9.69658261047591464109594-001	7.7492372047706947528014[illegible] 002
		8.43212729824415111461769-001	1.7376481812226193774613[illegible] 001
		6.24629009127993424428637-001	2.6446733948971187301823[illegible]-001
	K=4	9.7799271004812142431554[illegible] 001	5.6219952761090592877542[illegible] 002
		8.86255066359002048811827-001	1.2594395961866777519717[illegible] 001
		7.2951038131060032736884[illegible] 001	1.8568010288139639504502[illegible] 001
		5.1735251659797936429531[illegible] 001	2.3955643500880488979565[illegible] 001
M=15		A	C
	K=1	9.41176470588235294117647-001	1.5517226455097646228013[illegible] 001
	K=2	9.67072028519606879821428-001	3.4789573187217709796131[illegible] 002
		8.22401655690919235968046-001	2.1011935037272384927691[illegible] 001
	K=3	9.77771252255516879448507-001	5.6992214406360843394834[illegible] 002
		8.83134860905461539074876-001	1.3236540216275605294831[illegible] 001
		7.10492458267507010048045-001	2.1742546566669532099372[illegible] 001
	K=4	9.83591748700892685058383 001	4.2025996536087457621776[illegible] 002
		9.14231900362764409020027-001	9.6344354938421741128636[illegible] 002
		7.91621561625107369655687-001	1.4873057965416433535614[illegible] 001
		6.14902615398192058005035[illegible] 001	2.0843254317244653715431[illegible]-001

N=2		A	C
	K=1	9.5454545454545454545454545···e-1	1.2e7392369966411357e6291 e+1
	K=2	9.74297e-6234e-1e385339653···e-1	6.5422779815551244e117321 e+2
		8.59e-36327e-99322947993676···e-1	1.7e4e+2217817184e7794459···e+1
	K=3	9.8246e-5534582132464836l1 e+1	4.5e71319e+8275111817e1872···e+2
		9.e-6884533637172394387292e···e+1	1.e+344561284457263569221 e+1
		7.545e1e-6675e-7682e-5282945···e+1	1.8342294815798l2e9332793 e+1
	K=4	9.869193538213196l4e-6e8612···e+1	3.3555944257e318267e18985 e+2
		9.3115597728e-5629361e-4918 e+1	7.8ee5856702511526265572e e+2
		8.3e-536772ee-552496e32319 e+1	1.2378e+26365625ee186e199 e+1
		6.799593254629934936827l7···e-1	1.8217726ge-8345762966412 e+1

III.8. Effective Formulas

To calculate the integrals (6.2) by means of (6.3) we must evaluate $D_2 D^\alpha y_i$ at the point

(8.1) $$x_o + a_j h =: \xi_j \quad , \quad y(x_o + a_j h) =: \eta_j \ .$$

If we want to do this by means of the recursion formulas of Sec. II.6. we have to calculate the expressions $\left[Y_{i\alpha}\right]_{\xi_j, \eta_j}$ with the formulas of II.2. first, because these are needed for the general recursion formula. Then, the formulas of Sec. II.6. give the expressions (cf. (II.6.1))

$$\left[\overline{Y}_{i\nu}\right]_{\xi_j, \eta_j} = \left[\frac{D_2 D^\alpha y_i}{\alpha!}\right]_{\xi_j, \eta_j}$$

where (cf. (1.3)) iteration must be started with the values

$$\left[\overline{Y}_{io}\right]_{\xi_j, \eta_j} = \left[D_2 y_i\right]_{\xi_j, \eta_j} = f_i(\xi_j, \eta_j) - \hat{f}_i(\xi_j, \eta_j)$$

and where we have to put

$$\overline{X}_o = \overline{X}_1 = \ldots = 0$$

for the independent variable x and

$$\overline{C}_o = \overline{C}_1 = \ldots = 0$$

for a constant c.

Now, the formulas (6.2, 6.3) assume the form

(8.2)
$$\int_{x_o}^{x_o+h} \frac{(x_o+h-\xi)^\alpha}{\alpha!} \left[D_2 D^\alpha y_i\right]_{\xi, \hat{y}(\xi)} d\xi =$$

$$= h \sum_{j=1}^{k} c_j (x_o + h - x_o - a_j h)^\alpha \left[\overline{Y}_{i\alpha}\right]_{\xi_j, \eta_j} =$$

$$= h^{\alpha+1} \sum_{j=1}^{k} c_j (1-a_j)^\alpha \left[\overline{Y}_{i\alpha}\right]_{\xi_j, \eta_j} =$$

$$= h^{\alpha+1} \sum_{j=1}^{k} \gamma_{j\nu} \left[\overline{Y}_{i\alpha}\right]_{\xi_j, \eta_j}$$

where the quantities

$$\gamma_{i\alpha} = c_j(1-a_j)^{\alpha} \tag{8.3}$$

can be prepared at the very beginning.

III.9. Choosing the Orders m, s, and k

For choosing k, i.e., the number of the base points used in the quadrature formula (8.2), it is important to consider that the errors of the quadrature formulas and the methodical error caused by breaking off the series (2.1') should be of the same order of magnitude. Otherwise, it would make no sense going through the trouble of calculating higher terms of (2.1') while an error ten times as large has already been committed in the quadrature of the first (and usually largest) integral. On the other hand it makes also no sense to calculate the integrals with particular accuracy in view of a large breaking-off error. We shall therefore try to choose k in such a way that the quadrature formula is of at least the same order as the method, but then its order is not much higher either. As is known, Gauss's quadrature formula with k base points is of the order 2k. By means of the lemma in Sec. III.6 we succeeded to split off the factor t^m from the integrand (cf. (6.5)). Therefore, the order of the quadrature formula has been raised to m+2k. The method, on the other hand, is of the order m+s+1 (cf. Sec. III.5). Equating both orders we have

$$k \approx \frac{s+1}{2} \tag{9.1}$$

Hence, k should be about half as great as the number of integrals used.

The choice of m and s is a question of the differentiability properties of the differential equation as well as a question of expenditure. With the quoted recursion formulas, labor is approximately proportional to $(m+1)m+2k(s+1)(s+2)$ or, with (9.1), to $(m+1)m+(s+1)^2(s+2)$; thus, it increases with s much faster than with m. Minimizing this expression under the subsidiary condition of constant order m+s+1 one finds $(s+1)(3s+5)=2m+1$; this applies to the combinations

s	0	1	2	3	4
m	2	8	16	27	42

The choice of m and s is also a question of the desired accuracy. The influence of m and s on the results depends on the magnitude of the constants M and K_s of the theorem in Sec. III.5. A method of higher order is always better than a method of lower order, if the error limit is small enough. Usually, m is chosen between 5 and 20, s between 0 and 5. Then, with the limits of accuracy chosen, one tries to fit the step size along the solution.

III.10. Estimation of Error

The (methodical) error committed in one step may be estimated by means of the theorem of Sec. III.5 (e.g., by regarding the difference $\hat{y}(x)-\bar{y}(x)$ as the error in $\hat{y}(x)$) or by directly estimating the remainder formula (2.1"). The latter case will be considered here. We may replace (2.1") by

$$(10.1) \qquad R_{is}^{*} = \int_{x_o}^{x} \frac{(x-\xi)^s}{s!} \left\{ \left[D^{s+1} y_i \right]_{\xi,\bar{y}(\xi)} - \left[D^{s+1} y_i \right]_{\xi,\hat{y}(\xi)} \right\} d\xi \quad ,$$

for, owing to (2.1")

$$R_{is}^{**}(x) = \int_{x_o}^{x} \frac{(x-\xi)^s}{s!} \left\{ \left[D^{s+1} y_i \right]_{\xi,y(\xi)} - \left[D^{s+1} y_i \right]_{\xi,\bar{y}(\xi)} \right\} d\xi$$

is the error in the solution obtained from (2.1') when $\bar{y}(x)$ is used as a starting approximation. But owing to (5.2, 5.3), we have for this error

$$\left| R_{is}^{**}(x) \right| \leq M |x-x_o|^{m+1} \frac{\left[K_s n |x-x_o|^{s+1} \right]^2}{(m+1+2(s+1))!}$$

i.e., its order is much higher than that of R_{is}^{*}, therefore, it may be neglected.

The following lemma, which is about the order of the integrand

$$(10.2)\qquad z_i(x) := \frac{1}{s!}\left\{\left[D^{s+1}y_i\right]_{x,\bar{y}(x)} - \left[D^{s+1}y_i\right]_{x,\hat{y}(x)}\right\}$$

is useful to an expedient evaluation of (10.1).

Lemma: From the Lipschitz condition (5.1) for $D^s f_i, D^{s+1}y_i$ and from

$$y_i(x)-\hat{y}_i(x) = O((x-x_o)^{m+1})$$

it follows that

$$z_i(x) = O((x-x_o)^{m+1}) \qquad (i=1,\ldots,n).$$

Proof: From the theorem in Sec. III.5 follows $\bar{y}_i(x)-y_i(x)=O((x-x_o)^{m+s+2})$. This relation and $\bar{y}_i(x)-\hat{y}_i(x)=(\bar{y}_i(x)-y_i(x))+(y_i(x)-\hat{y}_i(x))=O((x-x_o)^{m+1})$ together with the Lipschitz condition give the statement.

Hence, we have

$$(10.3)\qquad z_i(x) = (x-x_o)^{m+1}Z_i(x) \quad .$$

With this expression we approximate the integral (10.1) by means of a quadrature formula which uses only the point x_o+h, i.e., the end point of the step of integration, as a base point. The function $(x-\xi)^s(\xi-x_o)^{m+1}$ is split off as a weight function. Therefore, we put

$$(10.4)\qquad \int_{x_o}^{x_o+h}(x_o+h-\xi)^s z_i(\xi)d\xi = \int_{x_o}^{x_o+h}(x_o+h-\xi)^s(\xi-x_o)^{m+1}Z_i(\xi)d\xi = cZ_i(x_o+h) \quad .$$

We determine the weight factor c in such a way that (10.4) is fulfilled exactly if $Z_i(\xi)$ is constant. This gives

$$c = \frac{s!(m+1)!}{(s+m+2)!}\,h^{m+s+2} \quad .$$

Inserting this weight factor in (10.4) we find the approximate error

$$(10.5)\qquad R_{is}(x) - R^*_{is}(x) = \int_{x_o}^{x_o+h}(x_o+h-\xi)^s z_i(\xi)d\xi =$$

$$= h^{s+1}\,\frac{s!(m+1)!}{(s+m+2)!}\,z_i(x_o+h) =$$

$$= h^{s+1}\,\frac{(m+1)!}{(s+m+2)!}\left\{\left[D^{s+1}y_i\right]_{x,\bar{y}(x)} - \left[D^{s+1}y_i\right]_{x,\hat{y}(x)}\right\} =$$

$$= h^{s+1}\,\frac{(s+1)!(m+1)!}{(s+m+2)!}\left\{\left[Y_{i,s+1}\right]_{x,\bar{y}(x)} - \left[Y_{i,s+1}\right]_{x,\hat{y}(x)}\right\}$$

(cf. (10.2), (II.2.3)) or, owing to (II.2.5)

$$R_{is}(x) = h^{s+1}\,\frac{s!(m+1)!}{(s+m+2)!}\left\{\left[F_{i,s+1}\right]_{x,\bar{y}(x)} - \left[F_{i,s+1}\right]_{x,\hat{y}(x)}\right\}. \tag{10.6}$$

III.11. Numerical Examples

With several simple examples having known solutions we studied the efficiency of formula (2.1') with (4.1),(4.1'),(8.2) and of the remainder (10.6) by means of the subroutines represented in Knapp-Wanner /29/ or Wanner /51/. In particular, we examined the question whether increasing s and simultaneously decreasing m, so that the total order m+s+1 remains constant, has a favorable effect on the result or not. In eleven out of twelve arbitrarily chosen examples, the result was positive, whereas only in one a higher number of Taylor terms turned out to be more expedient. Here are the results of the example

1) $\qquad y' = 1-e^{-y}(\sin x - \cos x)\ ,\ y(0) = 0$

with the solution $y(x) = \log(\sin x + e^x)$. The data given are the size h of the <u>single</u> step that was calculated, the orders m and s of the formulas (4.1),(2.1'), the actual errors of the Taylor series y(x) with m terms, the errors of the Lie-series solution (2.1'), and the estimate of the error given by the program according to (10.6):

h	m	s	Error in y	Error in $\bar{y}$	Estimates of error
0,125	18	0	$7,2 \cdot 10^{-15}$	$3,1 \cdot 10^{-17}$	$3,1 \cdot 10^{-17}$
	13	5	$2,3 \cdot 10^{-11}$	$1,45 \cdot 10^{-20}$	$1,0 \cdot 10^{-20}$
	8	10	$8,2 \cdot 10^{-8}$	$4,9 \cdot 10^{-21}$	$1,5 \cdot 10^{-21}$
0,250	18	0	$3,25 \cdot 10^{-9}$	$2,0 \cdot 10^{-11}$	$1,9 \cdot 10^{-11}$
	13		$3,19 \cdot 10^{-7}$	$6,6 \cdot 10^{-15}$	$3,3 \cdot 10^{-15}$
	8	10	$3,62 \cdot 10^{-5}$	$1,8 \cdot 10^{-15}$	$0,2 \cdot 10^{-15}$
0,500	18	0	$1,3 \cdot 10^{-3}$	$6,3 \cdot 10^{-6}$	$6,2 \cdot 10^{-6}$
	13	5	$4,1 \cdot 10^{-3}$	$1,1 \cdot 10^{-9}$	$4,7 \cdot 10^{-10}$
	8	10	$1,5 \cdot 10^{-2}$	$3,5 \cdot 10^{-10}$	$0,08 \cdot 10^{-10}$

Taylor's series, which converges only for $h \leq 0,5885...$ cannot be used for $h=0,5$ (also 0,25). Yet, the Lie-series correction gives good results. Estimation of the error is satisfactory, especially in the case of a reasonable step size and for the (usual) smaller values of s and greater values of m (cf. Sec. III.9).

2) Comparison with Runge-Kutta-Fehlberg:

Differential equations of restricted three body problem:

$$y_1' = y_2 \qquad (\mu' = 1-\mu)$$

$$y_2' = y_1 + 2y_4 - \mu' \frac{y_1+\mu}{((y_1+\mu)^2+y_3^2)^{3/2}} - \frac{y_1-\mu'}{((y_1-\mu')^2+y_3^2)^{3/2}}$$

$$y_3' = y_4$$

$$y_4' = y_3 - 2y_2 - \mu' \frac{y_3}{((y_1+\mu)^2+y_3^2)^{3/2}} - \frac{y_3}{((y_1-\mu')^2+y_3^2)^{3/2}} .$$

With this equation, in Durham /10/ a comparison of different methods was carried out and there the method of Fehlberg

turned out to be the best.
Using their initial values for three different Arenstorf orbits we reran these examples with our method, taking m=13, s=3.
For comparison the results are presented in the following table:

Table

The errors Δy_i represent the amount by which the initial conditions failed to be duplicated at the end T of the orbit (y_1-axis crossing) for the y_i coordinates respectively.
Fehlberg's results are taken from the above mentioned report.
Figures of the orbits and the exact initial datas can also be found in /51/, p.106-110.

orbit number	number of steps	positions Δy_1	positions Δy_3	velocities Δy_2	velocities Δy_4	method
1	269	0.3	not given	0.07	1	Fehlberg
	233	0.005	0.010	0.004	0.005	Lie-series
2	395	0.05	not given	0.1*)	1	Fehlberg
	219	0.009	0.028	4.635	0.381	Lie-series
3	284	0.1	not given	0.07*)	2	Fehlberg
	214	0.005	0.011	1.783	0.691	Lie-series

in units of 10^{-16}.

*) These values are probably not correct, since the connection matrix for orbit 2 after one period is

$$H(T) = \begin{pmatrix} 3.10*10^{3} & 6.03*10^{0} & -9.37*10^{2} & -1.91*10^{1} \\ 1.72*10^{6} & 3.14*10^{3} & -4.88*10^{5} & -1.06*10^{4} \\ 1.05*10^{4} & 1.91*10^{1} & -2.97*10^{3} & -6.50*10^{1} \\ 4.76*10^{5} & 9.25*10^{2} & -1.44*10^{5} & -2.93*10^{3} \end{pmatrix}$$

This has been calculated with using 6 terms in each step. Thus, the derivatives of $y_2(T)$ with respect to the initial values (second row) are dominant. During one period, <u>six</u> digits are lo...

3) Example for a boundary value problem:

$$y_1' = y_2$$
$$y_2' = \exp y_1 \qquad y_1(0) = y_1(1) = 0,\ y_2(0) = ?$$

Since y_{2o} is not known, we guess y_{2o}^* and calculate the corresponding trajectories $y_1^*(x)$, $y_2^*(x)$. If $y_1^*(1) \neq 0$ we correct y_{2o}^* with Newton's method

$$y_{2o} = y_{2o}^* - \frac{y_1^*(1)}{y_{12}(1)} \quad .$$

The convergence was as follows ($\gamma = 10^{-24}$, total time 5 seconds):

y_{2o}^*

-0.41

-0.46358

-0.46363259167

-0.46363259172426226173111149

-0.46363259172426226173113495.

This, of course, is a simple example only. Other examples for boundary value problems are carried out in /22/ p.73-94.
All computations were carried out in double precision (26D) on the CDC 3600 at the Mathematics Research Center, Madison Wisconsin.
Further Examples can be found in the Chapter on step size control (Ch. VI.).

Chapter IV

Gröbner's Integral Equation and Convergence Proofs

by G. Wanner and H. Reitberger

In this chapter we give a new proof of the integral equation of W. Gröbner. It is a generalization of the well-known "variation of constants formula" to nonlinear cases. It makes possible an easy approach to the formulas of the preceding chapter and to a number of further methods. It also leads to many iteration methods, for some of which we give convergence proofs.

Our thanks go to Prof.W. Gröbner, K.H. Kastlunger and K. Egle for their helpful discussions. We further wish to acknowledge the suggestions of Prof.W. Hahn, Graz.

IV.1 The Integral Equation of Groebner

In this equation appear derivatives of the solutions with respect to the initial values y_0. Therefore in this chapter the following changed notation is preferable:

We denote by $Y(X,x,y)$ resp. $\hat{Y}(X,x,y)$

the solutions of the differential equations (I.1.1) resp. (III.1.1), i.e.

$$(1.1)\quad \frac{\partial Y(X,x,y)}{\partial X} = f(X,Y(X,x,y)) \qquad \frac{\partial \hat{Y}(X,x,y)}{\partial X} = \hat{f}(X,\hat{Y}(X,x,y))$$

with the initial values x,y; thus with

$$(1.2)\quad Y(x,x,y) = y \qquad \hat{Y}(x,x,y) = y$$

This means, that the dependance of the solutions on the initial values x, y are now kept in mind. Specialization of these to the prescribed initial values x_0, y_0 leads to the functions of the preceding chapters

$$(1.3)\quad Y(x,x_0,y_0) = y(x), \qquad \hat{Y}(x,x_0,y_0) = \hat{y}(x)$$

The connection between the wanted solution $Y(X,x,y)$ and the approximate solution $\hat{Y}(X,x,y)$, which is assumed to be known, is given by the following theorem:

Theorem: *If* f, $\hat{f}$ *and* $\frac{\partial f(x,y)}{\partial y}$ *are continuous, then it holds that*

$$(1.4)\quad Y(X,x,y) = \hat{Y}(X,x,y) + \int_x^X \left[D_2\, Y(X,\xi,y)\right]_{\xi,\hat{Y}(\xi,x,y)}\, d\xi$$

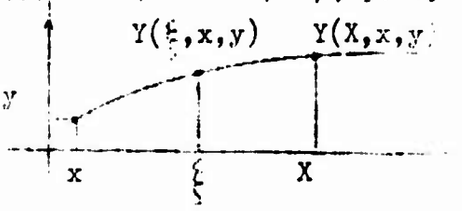
where

$$D_2 = \sum_i \left[f_i(x,y) - \hat{f}_i(x,y)\right] \frac{\partial}{\partial y_i} \quad .$$

Proof: From $\frac{\partial f}{\partial y} \in C$ it follows that $Y(X,x,y) \in C^1$ (cf. / 6/, p. 25)

We now differentiate the identity

$$Y(X,x,y) = Y(X,\xi,Y(\xi,x,y))$$



with respect to ξ and after that put $\xi = x$:

$$0 = \frac{\partial Y(X,x,y)}{\partial x} + f(x,y)\,\frac{\partial Y(X,x,y)}{\partial y}$$

This is possible because of $Y \in C^1$. Finally we insert ξ for x and $\hat{Y}(\xi,x,y)$ for y:

$$(1.5)\qquad 0 = \left[\frac{\partial Y(X,x,y)}{\partial x} + f(x,y)\,\frac{\partial Y(X,x,y)}{\partial y}\right]_{\xi,\hat{Y}(\xi,x,y)} .$$

A similar differentiation of $Y(X,\xi,\hat{Y}(\xi,x,y))$ with respect to ξ yields (using chain rule again)

$$\frac{\partial}{\partial \xi}\, Y(X,\xi,\hat{Y}(\xi,x,y)) = \left[\frac{\partial Y(X,x,y)}{\partial x} + \hat{f}(x,y)\,\frac{\partial Y(X,x,y)}{\partial y}\right]_{\xi,\hat{Y}(\xi,x,y)} .$$

Finally we subtract this from (1.5) and integrate from x to X:

$$\int_x^X \left[(f(x,y) - \hat{f}(x,y))\frac{\partial}{\partial y}\, Y(X,x,y)\right]_{\xi,\hat{Y}(\xi,x,y)} d\xi =$$

$$= \left[-Y(X,\xi,\hat{Y}(\xi,x,y))\right]_x^X = -\,Y(X,X,\hat{Y}(X,x,y)) + Y(X,x,\hat{Y}(x,x,y))$$

$$= -\hat{Y}(X,x,y) + Y(X,x,y) \qquad \text{(cf. (1.2))}.$$

Thus, (1.4) is proved. The different arguments ξ and x in $Y(X,\xi,y)$ do not mind, since they are equalised by the substitution rule $x \to \xi$, $y \to \hat{Y}(\xi,x,y)$.

Done.

This integral equation was found first in 1960 by Groebner for analytic equations. It was rediscovered in similar form (cf.(3.1)) in 1961 by **Alekseev** /54/ . The above given proof is similar to that of Alekseev.

IV.2 A Generalization

The above integral equation can be generalised in the following way:

Theorem: If f, $\hat{f}$, $\frac{\partial f}{\partial y}$ are continuous, $F(x,y)$ is continuously differentiable, it holds that

$$F(X,Y(X,x,y)) = F(X,\hat{Y}(X,x,y)) + \int_x^X \left[D_2 F(X,Y(X,\xi,y))\right]_{\xi,\hat{Y}(\xi,x,y)} d\xi \,. \tag{2.1}$$

Clearly (2.1) coincides with (1.4) if $F(x,y) \equiv y$. For analytic equations this formula has first been recognised by K. Egle (cf. QSR Nr.1).

Proof: First differentiate $F(X,Y(X,\xi,\hat{Y}(\xi,x,y)))$ with respect to ξ.

$$\frac{\partial}{\partial \xi} F(X,Y(X,\xi,\hat{Y}(\xi,x,y))) =$$
$$= \frac{\partial F}{\partial y}(X,Y(X,\xi,\hat{Y}(\xi,x,y))) \cdot \left[\frac{\partial Y(X,x,y)}{\partial x} + \hat{f}(x,y)\,\frac{\partial Y(X,x,y)}{\partial y}\right]_{\xi,\hat{Y}(\xi,x,y)}$$

Next we multiply (1.5) by $\frac{\partial}{\partial y} F(X,Y(X,\xi,\hat{Y}(\xi,x,y)))$ and subtract the two formulas:

$$-\frac{\partial}{\partial \xi} F(X,Y(X,\xi,\hat{Y}(\xi,x,y))) =$$
$$= \left[(f(x,y) - \hat{f}(x,y))\,\frac{\partial F}{\partial y}(X,Y(X,x,y))\,\frac{\partial Y(X,x,y)}{\partial y}\right]_{\xi,\hat{Y}(\xi,x,y)} =$$
$$= \left[(f(x,y) - \hat{f}(x,y))\,\frac{\partial}{\partial y} F(X,Y(X,x,y))\right]_{\xi,\hat{Y}(\xi,x,y)} .$$

Integration from x to X now yields the wanted integral equation (2.1), since again

$$\left[-F(X,Y(X,\xi,\hat{Y}(\xi,x,y)))\right]_x^X = -F(X,Y(X,X,\hat{Y}(X,x,y))) +$$
$$+ F(X,Y(X,x,\hat{Y}(x,x,y))) = -F(X,\hat{Y}(X,x,y)) + F(X,Y(X,x,y))$$

Done.

IV.3 A Volterra Integral Equation

Interchange Y and $\hat{Y}$ in (1.4):

$$\hat{Y} = Y + \int_x^X \underbrace{\left[(\hat{f} - f)\frac{\partial}{\partial y} Y\right]}_{-D_2} \Big|_{\xi, Y} d\xi$$

This yields the following Volterra integral equation for the solution Y.

Theorem: If f, $\hat{f}$ and $\frac{\partial \hat{f}(x,y)}{\partial y}$ are continuous, it holds that

$$(3.1)\quad Y(X,x,y) = \hat{Y}(X,x,y) + \int_x^X \left[D_2 \hat{Y}(X,\xi,y)\right]_{\xi, Y(\xi,x,y)} d\xi \ .$$

In addition, if F(x,y) is a continuously differentiable function, we have

$$(3.2)\quad F(X,Y(X,x,y)) = F(X,\hat{Y}(X,x,y)) + \int_x^X \left[D_2 F(X,\hat{Y}(X,\xi,y))\right]_{\xi, Y(\xi,x,y)} d\xi \ .$$

These formulas differ from (1.4) and (2.1) only by the exchange of Y and $\hat{Y}$ under the integral sign.

IV.4 The Variation of Constants-Formula as Special Case.

Let $\hat{f}(x,y)$ be linear in y:

$$\hat{Y}' = \hat{f}(X).\hat{Y}$$

and let $\hat{F}(x) = (\hat{F}_{ik}(x))$ be the fundamental system of solutions with $\hat{F}(x) = I$ (identity matrix). Then $\hat{Y}(X,\xi,y) = \hat{F}(X)\hat{F}^{-1}(\xi)y$, and $\frac{\partial}{\partial y}\hat{Y}(X,\xi,y) = \hat{F}(X)\,\hat{F}^{-1}(\xi)$. Thus for the solution of

$$Y' = \hat{f}(X)Y + s(X,Y(X))$$

(3.1) reads as follows:

$$(4.1)\quad Y(X,x,y) = \hat{F}(X)y + \int_{x}^{X} \hat{F}(\xi)\,\hat{F}^{-1}(\xi)\,g(\xi,Y(\xi,x,y))\,d\xi \quad .$$

This, however, is nothing else than the Variation of Constants-Formula for inhomogeneous linear differential systems.
This shows, that (3.1) acts the same part for nonlinear equations, than (4.1) does for linear equations. I.e., it has applications to asymptotic theory of differential equations (e.g. Wasow /52/,p.67, ... ; to stability theory (F. Brauer /55/, Alekseev /54/) or to the treatment of stiff differential equations. This we hope to discuss in a later report.

IV.5 Proof of the Formulas of Section III.2

The series of the preceding chapter are now simply obtained by inserting the Taylor series of the solution

$$(5.1)\quad Y(X,\xi,y) = \sum_{\alpha=0}^{s} \frac{(X-\xi)^{\alpha}}{\alpha!} D^{\alpha}y + \int_{\xi}^{X} \frac{(X-\eta)^{s}}{s!} \left[D^{s+1}y\right]_{\eta,Y(\eta,\xi,y)} d\eta$$

into the right hand side of (1.4):

$$(5.2)\quad Y(X,x,y) = \hat{Y}(X,x,y) + \sum_{\alpha=0}^{s} \int_{x}^{X} \frac{(X-\xi)^{\alpha}}{\alpha!} \left[D_2 D^{\alpha}y\right]_{\xi,\hat{Y}(\xi,x,y)} d\xi +$$

$$+ R_s(X,x,y)$$

with

$$R_s(X,x,y) = \int_{x}^{X} \int_{\xi}^{X} \left[D_2 \frac{(X-\eta)^{s}}{s!} \left[D^{s+1}y\right]_{\eta,Y(\eta,\xi,y)}\right]_{\xi,\hat{Y}(\xi,x,y)} d\eta\, d\xi$$

or by interchanging the order of integration:

$$R_s(X,x,y) = \int_{x}^{X} \int_{x}^{\eta} \left[D_2 \frac{(X-\eta)^{s}}{s!} \left[D^{s+1}y\right]_{\eta,Y(\eta,\xi,y)}\right]_{\xi,\hat{Y}(\xi,x,y)} d\xi\, d\eta \quad .$$

The inner integral of this can now be evaluated with the help of the generalized integral equation (2.1) to give

Theorem: If f, $\hat{f}$, $\frac{\partial f}{\partial y}$ are continuous and $D^{s+1}y$ is continuously differentiable, formula (5.2) is valid with the remainder

$$(5.3)\quad R_s(X,x,y) = \int_x^X \frac{(X-\eta)^s}{s!} \left\{ \left[D^{s+1}y\right]_{\eta,Y(\eta,x,y)} - \left[D^{s+1}y\right]_{\eta,\hat{Y}(\eta,x,y)} \right\} d\eta \ .$$

Proof: Under the given conditions (5.1) and (1.4) are valid. (2.1) is used with the function

$$F(x,y) = \frac{(X-x)^s}{s!} D^{s+1}y$$

and with X replaced by η. The stated conditions allow this application. Done.

(5.2) and (5.3) are nothing else than the formulas of III.3, if transcribed to the original notation.

Remark: Knapp,/26/, has proved (5.3) under the weaker condition $D^{s+1}y \in C$.

Following we prove the formula of Groebner in III.1:

IV.6 Convergence for $s \to \infty$

Theorem: If the functions $f_i(x,y)$, i.e., the operator D, are analytic in some domain, then for a sufficiently small $h=X-x$ (5.3) converges to zero, i.e. we have from (5.2)

$$(6.1)\quad Y(X,x,y) = \hat{Y}(X,x,y) + \sum_{\alpha=0}^{\infty} \int_x^X \frac{(X-\xi)^\alpha}{\alpha!} \left[D_2 D^\alpha y\right]_{\xi,\hat{Y}(\xi,x,y)} d\xi \ ,$$

the formula of Groebner stated in III.1.

Proof: Since D is analytic, it can be majorized by the operator Δ which is in one variable only

$$D \prec \Delta = \frac{N}{(1-\frac{z}{\varrho})} \frac{d}{dz}$$

(cf. e.g. Groebner /22/, p. 30, and Groebner-Watzlawek /22/, p. 225).

Thus

$$(6.2) \qquad |D^{s+1}y| \le \Delta^{s+1}z = \frac{(2s-3)(2s-5)\ldots 1.N^s}{\rho^s(1-z/\rho)^{2s-1}} \quad .$$

Next we shift the initial values y to the origin and choose h=X-x so that

$$(6.3) \qquad |Y(\xi)| \le \frac{\rho}{2}\ , |\hat{Y}(\xi)| \le \frac{\rho}{2} \qquad \text{for } x \le \xi \le X.$$

Inserting (6.2) into (5.3) we obtain

$$|R_s| \le \int_x^X \frac{(X-\xi)^s}{s!} \underbrace{(2s-3)(2s-5)\ldots}_{\le 2s(2s-2)(2s-4)\ldots \le 2^s.s!} \frac{N^s}{\rho^{s-1}} \left\{ \underbrace{\left|\frac{1}{(1-\frac{Y}{\rho})^{2s-1}}\right|}_{\le 2^{2s-1}} + \underbrace{\left|\frac{1}{(1-\frac{\hat{Y}}{\rho})^{2s-1}}\right|}_{\le 2^{2s-1}} \right\} d\xi \qquad (6.2)$$

$\le 2^{2s}$

Hence

$$|R_s| \le 2^{3s} \frac{N^s}{\rho^{s-1}} \frac{(X-x)^{s+1}}{s+1} = \frac{\rho^2}{8N(s+1)} \left[\frac{8N(X-x)}{\rho}\right]^{s+1}$$

and thus

$$\lim_{s\to\infty} R_s = 0$$

if

$$\frac{8N(X-x)}{\rho} \le 1 \qquad \text{or} \qquad h = |X-x| \le \frac{\rho}{8N} \ .$$

Done.

Remarks:

1) A second proof of (6.2) is possible by inserting the infinite power series into the integral and assuring uniform convergence which allows interchange of summation, integration and differentiation.

2) Still another proof (the historically first one) was given by Gröbner by rearranging the power series for the solution Y in a special way (cf. e.g. Gröbner /22/,p.35, Knapp-Wanner /29/,p.29).

IV.7. A General Process

Formula (5.2) has resulted from inserting a Taylor series solution into the integral of Gröbner's integral equation (1.4). This leads to the idea to insert any approximate solution $\tilde{Y}(X,\xi,y)$, say, of order s. We thus obtain a new approximate solution $\bar{Y}(X,x,y)$ given by the formula

$$(7.1)\qquad \bar{Y}(X,x,y) = \hat{Y}(X,x,y) + \int_x^X [D_2\tilde{Y}(X,\xi,y)]_{\xi,\hat{Y}(\xi,x,y)}\,d\xi\ .$$

Theorem. If $\hat{Y}$ is of order m, $\tilde{Y}$ is of order s, then, under appropriate differentiability conditions, $\bar{Y}$ is of order m+s+1.

Proof: Because of the order-condition, the error of $\tilde{Y}$ is equal to

$$\text{error of } \tilde{Y}(X,\xi,y) = \frac{(X-\xi)^{s+1}}{(s+1)!}\,F(X,\xi,y).$$

The error of $\bar{Y}$ is now obtained by subtraction of (7.1) from (1.4)

$$\text{error of } \bar{Y}(X,x,y) = \int_x^X [D_2(\text{error of } \tilde{Y}(X,\xi,y))]_{\xi,\hat{Y}(\xi,x,y)}\,d\xi.$$

Again, as in section III.6, D_2 contains the factor $(\xi-x)^m$ and we thus have

$$\text{error of } \bar{Y}(X,x,y) = \int_x^X \frac{(X-\xi)^{s+1}}{(s+1)!}\,\frac{(\xi-x)^m}{m!}\,G(\xi,x,y)\left[\frac{\partial}{\partial y}F(X,\xi,y)\right]_{\xi,\hat{Y}}\,d\xi.$$

Since $(X-\xi)^{s+1}(\xi-x)^m$ does not change sign in the integration interval, the mean value theorem can be applied and yields

$$\text{error } \bar{Y}(X,x,y) = \int_x^X \frac{(X-\xi)^{s+1}}{(s+1)!}\,\frac{(\xi-x)^m}{m!}\,d\xi\ G(\theta,x,y)\left[\frac{\partial}{\partial y}F(X,\theta,y)\right]_{\theta,\hat{Y}}$$

$$\frac{(X-x)^{m+s+2}}{(m+s+2)!} \qquad (x\le\theta\le X)$$

thus, $\bar{Y}$ is of order m+s+1. Done.

Remark: The Theorem in section III.5 is related to this.

Hence, to each pair of methods with orders m and s and with solutions $\hat{Y}$ and $\tilde{Y}$ resp., formula (7.1) leads to a new method with order m+s+1 and solution $\bar{Y}$. The Lie series of Chapter III is obtained by inserting m and s terms of the power series expansion.

IV.8. Iterated Integral Equations

Still further integral equations are derived from (1.4) and (2.1) by iteration:

$$(8.1)\quad Y(X,\xi_o,y) = \hat{Y}(X,\xi_o,y) + \int_{\xi_o}^{X} [D_2\hat{Y}(X,\xi_1,y)]_{\hat{z}_1} d\xi_1 + \int_{\xi_o}^{X}\int_{\xi_1}^{X} [D_2[D_2Y(X,\xi_2,y)]_{\hat{z}_2}]_{\hat{z}_1} d\xi_2 d\xi_1 \ .$$

and

$$(8.2)\quad F(X,Y(X,\xi_o,y)) = F(X,\hat{Y}(X,\xi_o,y)) + \int_{\xi_o}^{X} [D_2F(X,\hat{Y}(X,\xi_1,y))]_{\hat{z}_1} d\xi_1 + \int_{\xi_o}^{X}\int_{\xi_1}^{X} [D_2[D_2F(X,Y(X,\xi_2,y))]_{\hat{z}_2}]_{\hat{z}_1} d\xi_2 d\xi_1$$

where $\hat{z}_1, \hat{z}_2, \ldots$ denote the following substitution rule

$$(8.3)\qquad \hat{z}_k = x \to \xi_k, y \to \hat{Y}(\xi_k, \xi_{k-1}, y).$$

Still more general equations are derived after repeated iterations.
The following theorem is obtained by the repeated application of the preceding theorem (section IV.7):

Theorem: If $\hat{Y}$ is of order m, $\tilde{Y}$ is of order s, then, under appropriate differentiability conditions, $\bar{Y}$ which is defined by

$$(8.4)\quad \bar{Y}(X,\xi_o,y) = \hat{Y}(X,\xi_o,y) + \int_{\xi_o}^{X} [D_2\hat{Y}(X,\xi_1,y)]_{\hat{z}_1}\, d\xi_1 +$$

$$+ \int_{\xi_o}^{X}\int_{\xi_1}^{X} [D_2[D_2\tilde{Y}(X,\xi_2,y)]_{\hat{z}_2}]_{\hat{z}_1}\, d\xi_2 d\xi_1$$

is of order 2m+s+2. Done.

Remarks: 1) The insertion of a power series for Y leads to the series with multiple integrals as given in Wanner /51/,p.73-74

2) The order of the analoguous formula with an r-fold integral is rm+s+r.

IV.9. Iteration methods and convergence proofs

The formulas (7.1),(8.4) etc. can be iterated in many ways. There are possible iterations with respect to $\hat{Y}$, or with respect to $\tilde{Y}$, or both. A number of methods appear as special cases, such as the iteration methods of Picard or that of Gröbner-Knapp, the method of Poincaré and so on. For a few of these iteration methods convergence proofs and error estimates now are given, for others we have not yet found them.

IV.10. The Iteration Method of Gröbner-Knapp

This method appears, when (7.1) is iterated with respect to $\hat{Y}$ while for $\tilde{Y}$ the first s terms of the power series solution are inserted. Thus we have the iteration process (cf.(5.2))

(10.1)

$$Y^{(r+1)}(X,x,y) = Y^{(r)}(X,x,y) + \sum_{\alpha=0}^{s} \int_{x}^{X} \frac{(X-\xi)^{\alpha}}{\alpha!} [D_2^{(r)} D^{\alpha} y]_{\xi, Y^{(r)}(\xi,x,y)}\, d\xi$$

where

$$Y^{(r)\prime} = \hat{f}^{(r)}(x,Y^{(r)}),\quad D_2^{(r)} = \frac{\partial}{\partial x} + \hat{f}^{(r)}(x,y)\frac{\partial}{\partial y};$$

starting with $Y^{(o)} = \hat{Y}$.

The convergence follows from equation (III.5.3) under the condition that $D^{s+1}y$ satisfies a Lipschitzcondition. If the Lipschitz-condition is assured only in some compact domain B (as usual with nonlinear equations), further considerations are necessary to assure that the iterated functions $Y^{(r)}$ do not leave B:

Theorem: Assume that the Lipschitzcondition (III.5.1) for the functions $D^{s+1}y$ is satisfied in the domain

$$(10.2)\qquad B = \{(\xi,\eta)\,|\,x\le\xi\le x+a,\ |y_i-\eta_i|\le b\}$$

and that the approximate solution $\hat{Y}$ satisfies

$$(10.3)\qquad |Y_i(\xi,x,y) - \hat{Y}_i(\xi,x,y)| \le M\,\frac{|\xi-x|^{m+1}}{(m+1)!}$$

for $x\le\xi\le x+a$. Further $h=X-x$ should satisfy

$$(10.4)\qquad M\,\frac{h^{m+1}}{(m+1)!} \le \frac{b}{2}$$

$$(10.5)\qquad |Y_i(\xi,x,y) - y_i| \le \frac{b}{2} \qquad x\le\xi\le x+h$$

$$(10.6)\qquad \frac{K_s n h^{s+1}}{(m+2)(m+3)\ldots(m+s+2)} \le 1\,.$$

Then the iterated solutions $Y^{(r)}(\xi,x,y)$ of (10.1) do not leave B for $x\le\xi\le x+h$ and the iterations are arbitrary often possible and converge to the solution with the error estimation

$$(10.7)\qquad |Y_i(\xi,x,y) - Y_i^{(r)}(\xi,x,y)| \le M|\xi-x|^{m+1}\,\frac{(K_s n|\xi-x|^{s+1})^r}{(m+1+(s+1)r)!}\,.$$

Proof: 1) First we show that the functions $Y^{(r)}$ are again $(s+1)$-times differentiable, since only for $f,\hat{f}\in C^s$ the

formulas (5.2),(5.3) are valid. This we simply show by (s+1)-times differentiating R_s in (5.3):
Using the well known formula

$$\frac{d^{s+1}}{dX^{s+1}} \int_x^X \frac{(X-\xi)^s}{s!} g(\xi)d\xi = g(X)$$

we obtain

$$\frac{d^{s+1}}{dX^{s+1}} R_s(X,x,y) = [D^{s+1}y]_{X,Y} - [D^{s+1}y]_{X,\hat{Y}}$$

which is continuous. Now the assertion follows from the fact that $Y(X,x,y) \in C^{s+1}$ (cf. section II.1).

2) Next we confirm that the iterated functions do not leave B:
First by (III.5.3),(10.4),(10.6)

$$(10.8)\quad |Y_i - Y_i^{(1)}| \le \underbrace{\frac{M|\xi-x|^{m+1}}{(m+1)!}}_{\le \frac{b}{2}} \underbrace{\frac{K_s n|\xi-x|^{s+1}}{(m+2)..(m+s+2)}}_{\le 1} \le \frac{b}{2};$$

again by (III.5.3)

$$(10.9)\quad |Y_i - Y_i^{(2)}| \le \underbrace{\frac{M|\xi-x|^{m+1}}{(m+1)!}}_{\le \frac{b}{2}} \underbrace{\frac{K_s n|\xi-x|^{s+1}}{(m+2)..(m+s+2)}}_{\le 1} \underbrace{\frac{K_s n|\xi-x|^{s+1}}{(m+s+3)..(m+2s+3)}}_{<1} < \frac{b}{2}$$

and so on.
Thus by (10.5) and the triangle inequality

$$|y_i - Y_i^{(r)}| \le \underbrace{|y_i - Y_i|}_{\le \frac{b}{2}} + \underbrace{|Y_i - Y_i^{(r)}|}_{\le \frac{b}{2}} \le b.$$

(10.7) follows by induction from (10.8),(10.9),.. by (III.5.3).

Done.

Remarks. 1) The number m in (10.3),(10.4) need not be the greatest possible. Perhaps sometimes (10.4) may be less restrictive for smaller m; (10.6) however is not.

2) The theorem is essentially due to H.Knapp, the proof is new.

3) We already have a convergence proof for an arbitrary $\tilde{Y}$, not only for finite sections of a power series, and we are at the present working on its simplification.

IV.11. Picard's Method as Special Case

The method of Picard comes out from (10.1) by putting s=0:

$$Y^{(r+1)}(X,x,y) = Y^{(r)}(X,x,y) + \int_x^X [D_2^{(r)}y]_{\xi,Y^{(r)}}d\xi$$

$$\int_x^X f(\xi,Y^{(r)})d\xi - \underbrace{\int_x^X f^{(r)}(\xi,Y^{(r)})d\xi}_{-Y^{(r)}(X)+y}$$

$$= y + \int_x^X f(\xi,Y^{(r)}(\xi,x,y))d\xi,$$

which is the well-known iteration of Picard.,

IV.12. Poincare's Method of Parameter Expansion

If (7.1) iterated with respect to $\tilde{Y}$ while $\hat{Y}$ is kept fix, we get Poincaré's method of parameter expansion; i.e., we obtain the solution expanded in powers of a small parameter.
To show this, assume that the operator D_2 is multiplied with a small parameter, say ε:

$$D_2 = \varepsilon D_2^* .$$

$$(12.1)\quad Y^{(r+1)}(X,x,y) = \hat{Y}(X,x,y) + \int_x^X [D_2^* Y^{(r)}(X,\xi,y)]_{\xi,\hat{Y}}\, d\xi$$

starting, say, with $Y^{(o)}=\hat{Y}$. Inserting for $Y^{(r)}$ again and again, it can be seen, that Y is expanded in powers of ε(in simplified notation):

$$(12.2)\quad Y^{(r+1)} = \hat{Y} + \varepsilon\int D_2^*\hat{Y} + \varepsilon^2\int\int D_2^* D_2^*\hat{Y} + \ldots + \varepsilon^{r+1}\underbrace{\int\ldots\int}_{r+1}\underbrace{D_2^*\ldots D_2^*}_{r+1}\hat{Y}\ .$$

Subtracting this from (8.1) (more precisely: from the (r+1)-times iterated equation similar to (8.1)

$$(12.3)\quad Y = \hat{Y} + \varepsilon\int D_2^*\hat{Y} + \ldots + \varepsilon^{r+1}\underbrace{\int\ldots\int}_{r+1}\underbrace{D_2^*\ldots D_2^*}_{r+1}\hat{Y} + \varepsilon^{r+2}\underbrace{\int\ldots\int}_{r+2}\underbrace{D_2^*\ldots D_2^*}_{r+2}Y\ )$$

we obtain for the error

$$(12.4)\qquad Y - Y^{(r+1)} = \varepsilon^{r+2}\underbrace{\int\ldots\int}_{r+2}\underbrace{D_2^*\ldots D_2^*}_{r+2}Y\ .$$

We again consider the example

$$y' = \sqrt{x} + \sqrt{y},\ y(0)=0$$

which is known as an equation which poses difficulties to a numerical integration (cf. Rosser/42/, Cooper/8/). Again let

$$\hat{f} = \sqrt{x},\ D_2 = D_2^* = \sqrt{y}\,\frac{\partial}{\partial y}\ ,\ \varepsilon = 1.$$

Then the above expansion becomes

$$y(x) = \frac{2}{3}x^{3/2} + \sqrt{\frac{2}{3}}\frac{4}{7}x^{7/4} + \frac{1}{7}x^2 + \sqrt{\frac{2}{3}}\frac{1}{49}x^{9/4} + \ldots,$$

which, for small x, gives an exactitude of hundreds of Runge-Kutta steps.

A convergence proof of series (12.2) is given in section IV.14.

The actual determination of the series (12.2) is mostly easier by the usual expansion as e.g. it is described in Knapp-Wanner /28/, section IV.2..

IV.13. Power series as special case

Put

$$D_1 = \frac{\partial}{\partial x},\quad D_2 = \sum_i f_i(x,y)\,\frac{\partial}{\partial y_i},\quad \ldots = 1,\quad D_2^* = D_2$$

then

$$\hat{Y}(X,x,y) = y,$$

and the process (12.1) reads as follows

$$(13.1)\qquad Y^{(r+1)}(X,x,y) = y + \int_x^X f(\xi,y)\,\frac{\partial}{\partial y} Y^{(r)}(X,\xi,y)\,d\xi$$

a method, which is connected with a power series expansion in the following way:

Proposition: In the case of autonomous equations, (13.1) coincides with the method of power series:

This is seen by induction: For autonomous equations f(x,y) does not depend on x and we have for (13.1)

$$Y^{(r+1)} = y + \underbrace{f(y)\,\frac{\partial}{\partial y}}_{D} \int_x^X Y^{(r)}(X,\xi,y)\,d\xi.$$

Thus if

$$Y^{(r)} = \sum_{\alpha=0}^{r} \frac{(X-x)^\alpha}{\alpha!}\,D^\alpha y$$

then

$$Y^{(r+1)} = y + \sum_{\alpha=0}^{r} D \int_x^X \underbrace{\frac{(X-\xi)^\alpha}{\alpha!}}_{\frac{(X-x)^{\alpha+1}}{(\alpha+1)!}}\,d\xi\; D^\alpha y$$

and with $\beta=\alpha+1$

$$Y^{(r+1)} = \sum_{\beta=0}^{r+1} \frac{(X-x)^\beta}{\beta!}\,D^\beta y.$$

IV.14. Convergence proof of Poincaré's method

The following convergence proof of the iteration (12.1), starting with an arbitrary analytic function

$$Y^{(o)}(X,x,y) = \sum_{\nu=o}^{\infty} \frac{(X-x)^{\nu}}{\nu!} \bar{D}_1^{\nu} y$$

has been worked out together with K. Kuhnert.
First get a majorization operator

$$\Delta = \frac{N}{(1-z/\rho)} \frac{d}{dz}$$

for D, D_1, $\bar{D}_1$, as well as $\varepsilon D_2^* = D_2$.

Then

$$Y(X,x,y) - Y^{(r)}(X,x,y) \prec 2 \sum_{\alpha=r}^{\infty} \binom{\alpha}{r} \frac{(X-x)^{\alpha}}{\alpha!} \Delta^{\alpha} z. \tag{14.1}$$

where the symbol $\prec$ denotes majorization.
(14.1) is proved by induction:

$r=0$:

$$Y(X,x,y) - Y^{(o)}(X,x,y) = \sum_{\alpha=o}^{\infty} \frac{(X-x)^{\alpha}}{\alpha!} (D^{\alpha} z - \bar{D}_1^{\alpha} z)$$

$$\prec 2 \sum_{\alpha=o}^{\infty} \underbrace{\binom{\alpha}{o}}_{=1} \frac{(X-x)^{\alpha}}{\alpha!} \Delta^{\alpha} z .$$

$r \to r+1$:

$$Y(X,x,y) - Y^{(r+1)}(X,x,y) = \varepsilon \int_x^X [D_2^*(Y-Y^{(r)}]_{\hat{z}_1} \, d\xi$$

$$\prec \int_x^X [\Delta 2 \sum_{\alpha=r}^{\infty} \binom{\alpha}{r} \frac{(X-x)^{\alpha}}{\alpha!} \Delta^{\alpha} z]_{\xi, \sum_{\gamma=o}^{\infty} \frac{(\xi-x)^{\gamma}}{\gamma!} \Delta^{\gamma} z} \, d\xi .$$

Application of the commutation theorem (cf. e.g. /22/,p.17) and rearrangement of the double sum gives

$$Y-Y^{(r+1)} \prec 2 \sum_{\alpha=r}^{\infty} \sum_{\gamma=0}^{\infty} \binom{\alpha}{r} \underbrace{\int_X^X \frac{(X-\xi)^\alpha}{\alpha!} \frac{(\xi-x)^\gamma}{\gamma!} d\xi}_{\frac{(X-x)^{\alpha+\gamma+1}}{(\alpha+\gamma+1)!}} . \Delta^{\gamma+\alpha+1} z$$

and with $\beta=\alpha+1+\gamma$

$$Y-Y^{(r+1)} \prec 2 \sum_{\beta=r+1}^{\infty} \underbrace{\sum_{\alpha=r}^{\beta-1} \binom{\alpha}{r}}_{\binom{\beta}{r+1}} \frac{(X-x)^\beta}{\beta!} \Delta^\beta z$$

Done.

Next we transform the initial values y to the origine and sum up (14.1) for r=0,1,..:

$$\sum_{r=0}^{\infty} \sum_{\alpha=r}^{\infty} \binom{\alpha}{r} \frac{(X-x)^\alpha}{\alpha!} \Delta^\alpha z$$

$$= \sum_{\alpha=0}^{\infty} \underbrace{\sum_{r=0}^{\alpha} \binom{\alpha}{r}}_{2} \frac{(X-x)^\alpha}{\alpha!} \Delta^\alpha z$$

which converges for

$$|X-x| < \frac{\rho}{2N} .$$

Thus it is necessarily

$$\lim_{r\to\infty} (Y-Y^{(r)}) = 0.$$

Chapter V

Runge-Kutta Processes with Multiple Nodes

by K.H. Kastlunger

Abstract:

The use of the differential operator D makes it possible to extend the method of Runge-Kutta in such a way, that the power-series method, the classical Runge-Kutta-method as well as the processes of Fehlberg are contained as special cases. The generalization is in such a way, that not only the functions $f_i(x,y)$ are evaluated at some intermediate points, but in addition also the functions Df_i , D^2f_i , ... , D^mf_i . These new methods are advantageous especially when combined with the concept of recursive generation of the values of $D^nf_i = D^{n+1}y_i$, as desribed in Chapter II .

We first develop the general form of the conditions for the coefficients of these processes thereby extending the results of J.C. Butcher / 1/. These equations become still more complicated than those for classical processes.

Next it is shown that to each quadrature formula with multiple nodes there exists an analoguous Runge-Kutta process with the same number of nodes and with the same order. This again extends results of Butcher / 2/.

Fehlberg's method is shown to be nothing else than a generalized Runge-Kutta process with one m-fold first node and a few additional single ones.

Finally we give examples of explicite process and numerical examples.

V.1. General Theory

Notation

Let the following autonomous system of ordinary differential equations be given

(1.1) $y_i' = f_i(y_1,\dots,y_n)$ or for short $y' = f(y)$

with the initial conditions

(1.2) $y_i(x_0) = y_{i0}$ or $y(x_0) = y_0$.

Here we treat autonomous systems, since then the following theory becomes more simple. This, of course, is no reduction of generality, what can be seen by introducing $x = y_0$ as new variable with $y_0' = 1$.

Are the functions f_i analytic in a neighbourhood of y_{i0}, then the solution of (1.1,2) is given by the following power series

$$(1.3)\quad y(x) = \sum_{k=0}^{\infty} \frac{(x-x_0)^k}{k!} \left[D^k y\right]_0 = \sum_{k=0}^{\infty} \frac{h^k}{k!} \left[D^{k-1} f\right]_0$$

with the differential operator

$$(1.4)\quad D = \sum_{j=1}^{n} f_j \frac{\partial}{\partial y_j}$$

Again the symbol $\left[\dots\right]_0$ means that after all differentiations the initial values x_0, y_0 are to be inserted.

Elementary Differentials

Definition: (Butcher / 1 /, p.187)

$\{1\} := f$ is the only one elementary differential of order one;

$$(1.5)\quad F = \{F_1 \dots F_s\} := \sum_{j_1,\dots,j_s=1}^{n} F_{1j_1} \dots F_{sj_s} \frac{\partial^s}{\partial y_{j_1} \dots \partial y_{j_s}}$$

where $F_i = (F_{i1},\dots,F_{in})$

is an elementary differential of order r and degree s, if F_i are elementary differentials of order r_i :

order r: $r = r_1 + \dots + r_s + 1$

degree s: s = number of F_i, which constitute F .

more generally:

(1.6a) $F = \{F_1^{\mu_1} \dots F_\sigma^{\mu_\sigma}\} := \{\underbrace{F_1 \dots F_1}_{\mu_1} \dots \underbrace{F_\sigma \dots F_\sigma}_{\mu_\sigma}\}$

is a elementary differential of order r and degree s, where

(1.6b) $r = \mu_1 r_1 + \dots + \mu_\sigma r_\sigma + 1$ and $s = \mu_1 + \dots + \mu_\sigma$.

In this notation the exponent may be zero also ; we also define

$\{F_1^0 \dots F_\sigma^0\} := f$.

Notation convention:

For simplification we now introduce the following notation:

if $v_1, \dots, v_s$ are vectors $v_i = (v_{i1}, \dots, v_{in})$, then we denote:

(1.7) $$v_1 \dots v_s \frac{d^s}{dy^s} := \sum_{j_1, \dots, j_s = 1}^{n} v_{1j_1} \dots v_{sj_s} \frac{\partial^s}{\partial y_{j_1} \dots \partial y_{j_s}} ;$$

more generally: with

(1.8a) $$s = \sum_{i=1}^{\sigma} \mu_i$$

we write

(1.8b) $$v_1^{\mu_1} \dots v_\sigma^{\mu_\sigma} \frac{d^s}{dy^s} := \underbrace{v_1 \dots v_1}_{\mu_1} \cdot \dots \cdot \underbrace{v_\sigma \dots v_\sigma}_{\mu_\sigma} \frac{d^s}{dy^s} .$$

Using this notation (1.5) and (1.6) now becomes

(1.9) $$\{F_1 \dots F_s\} = F_1 \dots F_s \frac{d^s f}{dy^s} , \quad \{F_1^{\mu_1} \dots F_\sigma^{\mu_\sigma}\} = F_1^{\mu_1} \dots F_\sigma^{\mu_\sigma} \frac{d^s f}{dy^s}$$

The following theorem about elementary differentials is due to Butcher / 1 /

Theorem: $D^{r-1}f$ is a linear combination of all elementary differentials F of order r with positive integer coefficients α:

(1.10) $$D^{r-1}f = \sum_{\text{Ord } F = r} \alpha F \quad .$$ [1)]

The coefficient α of $F = \{F_1^{\mu_1} \dots F_\sigma^{\mu_\sigma}\}$ is given by

---

1) The symbol $\sum_{\text{Ord } F = r}$ denotes, that the summation is over all elementary differentials of order r.

(1.11) $$\alpha = (r-1)!\ \frac{\sigma}{\prod_{i=1}} \frac{1}{\mu_i!}\left(\frac{\alpha_i}{r_i!}\right)^{\mu_i} ,$$

where α_i are the coefficients of F_i, r_i their orders and r the order of F.

Examples:

$Df = \{f\}$

$D^2 f = \{\{f\}\} + \{f^2\}$

$D^3 f = \{\{\{f\}\}\} + 3\{\{f\}f\} + \{\{f^2\}\} + \{f^3\}$.

A list of all elementary differentials up to order eight is given in /1/, pp. 191-193.

We next modify the above theorem:

with $D^{r_i-1} f = D^{r_i} y = \sum_{\text{Ord } F_i = r_i} \alpha_i F_i$ we have

(1.12) $$\{(D^{r_1}y)\ldots(D^{r_s}y)\} = \sum_{\text{Ord } F_1 = r_1} \cdots \sum_{\text{Ord } F_s = r_s} \alpha_1 \cdots \alpha_s \{F_1 \ldots F_s\}$$

Theorem: It holds that

(1.13) $$D^r y = \sum_{t=1}^{r-1} \sum_{\substack{r_1+\ldots+r_t=r-1 \\ r_i \geq 1}} \frac{1}{t!} \cdot \frac{(r-1)!}{r_1!\ldots r_t!} \{(D^{r_1}y)\ldots(D^{r_t}y)\} .$$

Proof: Inserting (1.12) into (1.13) we obtain

$$\sum_{t=1}^{r-1} \sum_{\substack{r_1+\ldots+r_t=r-1 \\ r_i \geq 1}} \frac{(r-1)!}{r_1!\ldots r_t!} \cdot \frac{1}{t!} \sum_{\text{Ord } F_1 = r_1} \cdots \sum_{\text{Ord } F_t = r_t} \alpha_1 \ldots \alpha_t \{F_1 \ldots F_t\}$$

It is easy seen that all elementary differentials of order r appear in this expansion. If we pick out one of these, say $\{F_1^{\mu_1} \ldots F_\sigma^{\mu_\sigma}\}$, we see that the coefficient of this is equal to

$$\frac{(r-1)!}{t!} \frac{\sigma}{\prod_{j=1}} \left(\frac{\alpha_j}{r_j!}\right)^{\mu_j} \cdot \frac{t!}{\mu_1!\ldots\mu_\sigma!} = (r-1)! \frac{\sigma}{\prod_{j=1}} \frac{1}{\mu_j!}\left(\frac{\alpha_j}{r_j!}\right)^{\mu_j} .$$

This, however is exactly the coefficient α .

Power Series for the Runge-Kutta Approximation

The classical Runge-Kutta-method uses the following formulas for the approximation $\hat{y}(x)$:

$$\hat{y}(x) = y_0 + h \sum_{i=1}^{n} c_i g_i \tag{1.14a}$$

and

$$g_i = f(y_0 + h \sum_{j=1}^{n} b_{ij} g_j) \quad . \tag{1.14b}$$

These formulas are now generalized in the following way:

n = number of nodes(stages) of the method
m = their multiplicity

$$g_i^{(k)} = (D^k y)(y_0 + h \sum_{j=1}^{n} b_{ij}^{(1)} g_j^{(1)} + \frac{h^2}{2!} \sum_{j=1}^{n} b_{ij}^{(2)} g_j^{(2)} + \ldots + \frac{h^m}{m!} \sum_{j=1}^{n} b_{ij}^{(m)} g_j^{(m)}) \tag{1.15}$$

$$\hat{y}(x) = y_0 + h \sum_{i=1}^{n} c_i^{(1)} g_i^{(1)} + h^2 \sum_{i=1}^{n} c_i^{(2)} g_i^{(2)} + \ldots + h^m \sum_{i=1}^{n} c_i^{(m)} g_i^{(m)} \tag{1.16}$$

Remarks:

1) For m=1 we get (1.14) again.

2) For simplificity we put

$$\begin{aligned} c_i^{(k)} &= 0 \quad \text{for } k=m+1, m+2, \ldots \quad , i=1,2,\ldots,n \\ b_{ij}^{(k)} &= 0 \quad \text{for } k=m+1, m+2, \ldots \quad , i,j=1,2,\ldots,n \end{aligned} \tag{1.17}$$

3) The assumption that <u>all</u> nodes have the <u>same</u> multiplicity is not restrictive. Otherwise we put

$m = \max \{m_1, \ldots, m_n\}$, m_i = multiplicity of the i-th node

$$\begin{aligned} c_i^{(k)} &= 0 \quad \text{for } k= m_i+1, m_i+2, \ldots \quad , i=1,\ldots,n \\ b_{ij}^{(k)} &= 0 \quad \text{for } k=m_j+1, m_j+2, \ldots \quad , i,j=1,\ldots,n \end{aligned} \tag{1.18}$$

<u>Theorem</u> (<u>Expansion Theorem for Runge-Kutta</u>):

a) $g_i^{(k)}$ <u>posesses the following power series</u>

$$g_i^{(k)} = \sum_{\mu=0}^{\infty} \frac{h^\mu}{\mu!} R_{i,\mu}^{(k)} \quad ; \tag{1.19}$$

the vectors $R_{i,\mu}^{(k)}$ are determined recursivly:

$$(1.20a)\quad R_{i,0}^{(k)} = \left[D^k y\right]_0 \qquad k=1,2,\ldots$$

$$(1.20b)\quad R_{i,\mu}^{(k)} = \sum_{\tau=1}^{\mu} \frac{1}{\tau!} \sum_{\substack{\nu_1+\ldots+\nu_\tau=\mu \\ \nu_i \geq 1}} \frac{\mu!}{\nu_1!\ldots\nu_\tau!} S_{i,\nu_1}\cdots S_{i,\nu_\tau} \left[\frac{d^\tau}{dy^\tau} D^k y\right]_0$$

$$k=1,2,\ldots\ ,\ \mu=1,2,\ldots$$

$$(1.21)\quad S_{i,\nu} = \sum_{\tau=1}^{\nu} \binom{\nu}{\tau} \sum_{j=1}^{n} b_{ij}^{(\tau)} R_{j,\nu-\tau}^{(\tau)} \qquad \nu=1,2,\ldots$$

b) The approximation of the Runge-Kutta-method has the following expansion:

$$(1.22)\quad \hat{y}(x) = y_0 + \sum_{\nu=1}^{\infty} \frac{h^\nu}{\nu!} T_\nu$$

$$(1.23)\quad T_\nu = \sum_{k=1}^{\nu} \frac{\nu!}{(\nu-k)!} \sum_{i=1}^{n} c_i^{(k)} R_{i,\nu-k}^{(k)} \quad .$$

Proof:

1) Since f is analytic, it follows that $g_i^{(k)}$ posesses a power-series expansion (1.19) .

2) $y_0 + \sum_{\tau=1}^{n} \frac{h^\tau}{\tau!} \sum_{j=1}^{n} b_{ij}^{(\tau)} g_j^{(\tau)}$ (1.17) $y_0 + \sum_{\tau=1}^{\infty} \frac{h^\tau}{\tau!} \sum_{j=1}^{n} b_{ij}^{(\tau)} g_j^{(\tau)}$ (1.19)

$$y_0 + \sum_{\tau=1}^{\infty} \frac{h^\tau}{\tau!} \sum_{j=1}^{n} b_{ij}^{(\tau)} \sum_{\mu=0}^{\infty} \frac{h^\mu}{\mu!} R_{j,\mu}^{(\tau)} =$$

$$y_0 + \sum_{\tau=1}^{\infty} \sum_{\mu=0}^{\infty} \frac{h^{\tau+\mu}}{(\nu+\mu)!} \cdot \frac{(\nu+\mu)!}{\tau!\mu!} \sum_{j=1}^{n} b_{ij}^{(\tau)} R_{j,\mu}^{(\tau)} \qquad (\nu+\mu=\nu)$$

$$y_0 + \sum_{\nu=1}^{\infty} \frac{h^\nu}{\nu!} \sum_{\tau=1}^{\nu} \binom{\nu}{\tau} \sum_{j=1}^{n} b_{ij}^{(\tau)} R_{j,\nu-\tau}^{(\tau)} = y_0 + \sum_{\nu=1}^{\infty} \frac{h^\nu}{\nu!} S_{i,\nu} = y_0 + v_i$$

with

$$(1.24)\quad v_i = \sum_{\nu=1}^{\infty} \frac{h^\nu}{\nu!} S_{i,\nu}$$

$$(1.21)\quad S_{i,\nu} = \sum_{\tau=1}^{\nu} \binom{\nu}{\tau} \sum_{j=1}^{n} b_{ij}^{(\tau)} R_{j,\nu-\tau}^{(\tau)}$$

3) $$v_i^\tau \frac{d^\tau}{dy^\tau} = \left(\sum_{\nu=1}^{\infty} \frac{h^\nu}{\nu!} S_{i,\nu}\right)^\tau \frac{d^\tau}{dy^\tau} =$$

$$= \sum_{\varkappa_1=1}^{\infty} \cdots \sum_{\varkappa_\tau=1}^{\infty} \frac{h^{\varkappa_1+\ldots+\varkappa_\tau}}{\varkappa_1!\ldots\varkappa_\tau!} S_{i,\varkappa_1} \cdots S_{i,\varkappa_\tau} \frac{d^\tau}{dy^\tau} \quad (\varkappa_1+\ldots+\varkappa_\tau=\mu)$$

$$(1.25)\quad \sum_{\mu=\tau}^{\infty} \frac{h^\mu}{\mu!} \sum_{\substack{\varkappa_1+\ldots+\varkappa_\tau=\mu \\ \varkappa_i \geqslant 1}} \frac{\mu!}{\varkappa_1!\ldots\varkappa_\tau!} S_{i,\varkappa_1} \cdots S_{i,\varkappa_\tau} \frac{d^\tau}{dy^\tau}$$

4) Taylor's theorem for multiple variables with use of (1.7) and (1.8):

$$(1.26)\quad f(y_o+v) = \sum_{\tau=0}^{\infty} \frac{1}{\tau!} v^\tau \left[\frac{d^\tau}{dy^\tau} f\right]_o$$

5) $$g_i^{(k)} = (D^k y)(y_o+v_i) \quad (1.26)$$

$$\left[D^k y\right]_o + \sum_{\tau=1}^{\infty} \frac{1}{\tau!} v_i^\tau \left[\frac{d^\tau}{dy^\tau} D^k y\right]_o \quad (1.25)$$

$$\left[D^k y\right]_o + \sum_{\tau=1}^{\infty} \frac{1}{\tau!} \sum_{\mu=\tau}^{\infty} \frac{h^\mu}{\mu!} \sum_{\substack{\varkappa_1+\ldots+\varkappa_\tau=\mu \\ \varkappa_i \geqslant 1}} \frac{\mu!}{\varkappa_1!\ldots\varkappa_\tau!} S_{i,\varkappa_1} \cdots S_{i,\varkappa_\tau} \left[\frac{d^\tau}{dy^\tau} D^k y\right]_o =$$

$$\left[D^k y\right]_o + \sum_{\mu=1}^{\infty} \frac{h^\mu}{\mu!} \sum_{\tau=1}^{\mu} \frac{1}{\tau!} \sum_{\substack{\varkappa_1+\ldots+\varkappa_\tau=\mu \\ \varkappa_i \geqslant 1}} \frac{\mu!}{\varkappa_1!\ldots\varkappa_\tau!} S_{i,\varkappa_1} \cdots S_{i,\varkappa_\tau} \left[\frac{d^\tau}{dy^\tau} D^k y\right]_o .$$

A comparison with (1.19) gives

$$(1.2oa)\quad R_{i,o}^{(k)} = \left[D^k y\right]_o$$

$$(1.2ob)\quad R_{i,\mu}^{(k)} = \sum_{\tau=1}^{\infty} \frac{1}{\tau!} \sum_{\substack{\varkappa_1+\ldots+\varkappa_\tau=\mu \\ \varkappa_i \geqslant 1}} \frac{\mu!}{\varkappa_1!\ldots\varkappa_\tau!} S_{i,\varkappa_1} \cdots S_{i,\varkappa_\tau} \left[\frac{d^\tau}{dy^\tau} D^k y\right]_o$$

6) $$y(x) = y_o + h \sum_{i=1}^{n} c_i^{(1)} g_i^{(1)} + \ldots + h^m \sum_{i=1}^{n} c_i^{(m)} g_i^{(m)} \quad (1.17)$$

$$y_o + \sum_{k=1}^{\infty} h^k \sum_{i=1}^{n} c_i^{(k)} g_i^{(k)} \quad (1.19)$$

$$y_o + \sum_{k=1}^{\infty} h^k \sum_{i=1}^{n} c_i^{(k)} \sum_{\mu=o}^{\infty} \frac{h^\mu}{\mu!} R_{i,\mu}^{(k)} =$$

$$y_o + \sum_{k=1}^{\infty} \sum_{\mu=o}^{\infty} \frac{h^{k+\mu}}{(k+\mu)!} \cdot \frac{(k+\mu)!}{\mu!} \sum_{i=1}^{n} c_i^{(k)} R_{i,\mu}^{(k)} \quad (k+\mu=\varkappa)$$

$$y_o + \sum_{\varkappa=1}^{\infty} \frac{h^\varkappa}{\varkappa!} \sum_{k=1}^{\varkappa} \frac{\varkappa!}{(\varkappa-k)!} \sum_{i=1}^{n} c_i^{(k)} R_{i,\varkappa-k}^{(k)} = y_o + \sum_{\varkappa=1}^{\infty} \frac{h^\varkappa}{\varkappa!} T_\varkappa$$

with

$$(1.23)\quad T_\nu = \sum_{k=1}^{\nu} \frac{\nu!}{(\nu-k)!} \sum_{i=1}^{n} c_i^{(k)} R_{i,\nu-k}^{(k)} \quad . \qquad \text{Done.}$$

Connection with Elementary Differentials

Theorem: T_ν is a linear combination of all elementary differentials of order ν with coefficients $\beta.\phi$:

$$(1.27)\quad T_\nu = \sum_{k=1}^{\nu} \frac{\nu!}{(\nu-k)!} \sum_{i=1}^{n} c_i^{(k)} R_{i,\nu-k}^{(k)} = \sum_{\text{ord } F=\nu} \beta\phi [F]_o$$

where $\beta \in \mathbb{N}$ and ϕ has the following form

$$(1.28)\quad \phi = \phi^{(1)} + \ldots + \phi^{(\nu)} = \sum_{i=1}^{n} c_i^{(1)} \psi_i^{(1)} + \ldots + \sum_{i=1}^{n} c_i^{(\nu)} \psi_i^{(\nu)}$$

$$(1.29)\quad \text{with} \quad \phi^{(k)} = \sum_{i=1}^{n} c_i^{(k)} \psi_i^{(k)} \qquad k=1,\ldots,\nu \ ;$$

$\psi_i^{(k)}$ are polynomials in $b_{ij}^{(k)}$ over $\mathbb{Q}$.

Proof: First we show that

$$(1.30)\quad R_{i,\nu-k}^{(k)} = \sum_{\text{ord } F=\nu} \beta' \psi_i^{(k)} [F]_o \qquad k=1,\ldots,\nu \ ; \ i=1,\ldots,n \ ; \ \nu=1,2,\ldots$$

Now the right hand side of (1.27) is proved by inserting (1.30) into the left hand side of (1.27) using equation (1.28).

The proof of (1.30) is by induction on ν:

$\nu=1$: here $k=1$ and

$$R_{i,\cdot}^{(1)} \ (1.20a) \ [D^1 y]_o = [f]_\cdot \ ;$$

induction from $\nu-1$ to ν:

let

$$(1.31)\quad R_{i,\mu-l}^{(l)} = \sum_{\text{ord } F=\mu} \beta' \psi_i^{(l)} [F]_o \qquad (\ l=1,\ldots,\mu \ ; \ \mu=1,\ldots,\nu-1 \ ) \ ;$$

then

$$R_{i,\nu-k}^{(k)} \quad (1.20b)$$

$$\sum_{\tau=1}^{\nu-k} \frac{1}{\tau!} \sum_{\substack{\nu_1+\ldots+\nu_\tau=\nu-k \\ \nu_1 \geq 1}} \frac{(\nu-k)!}{\nu_1!\ldots\nu_\tau!} S_{i,\nu_1} \ldots S_{i,\nu_\tau} \left[\frac{d^\tau}{dy^\tau} D^k y\right]_o \quad (1.21)$$

$$(1.32)\quad = \sum_{\tau=1}^{\kappa-k} \sum_{\substack{\kappa_1+\ldots+\kappa_\tau=\kappa-k \\ \kappa_1 \geq 1}} \sum_{\sigma_1=1}^{\kappa_1} \cdots \sum_{\sigma_\tau=1}^{\kappa_\tau} \left(\frac{1}{\tau!} \frac{(\kappa-k)!}{\kappa_1! \ldots \kappa_\tau!} \binom{\kappa_1}{\sigma_1} \cdots \binom{\kappa_\tau}{\sigma_\tau} \right) \cdot \left(\sum_{j_1=1}^{n} \cdots \sum_{j_\tau=1}^{n} b_{ij_1}^{(\sigma_1)} \cdots b_{ij_\tau}^{(\sigma_\tau)} \right) \cdot R_{j_1,\kappa_1-\sigma_1}^{(\sigma_1)} \cdots R_{j_\tau,\kappa_\tau-\sigma_\tau}^{(\sigma_\kappa)} \left[\frac{d^\tau}{dy^\tau} D^k y \right]_o$$

Inserting here the induction hypothesis we obtain, after some computation, a linear combination of expressions of the form $\left[F_1 \ldots F_\tau \frac{d^\tau}{dy^\tau} D^k y\right]_o$, where F_i is of order κ_i.

In Lemma 2 (p.80) we shall prove, that the above expression is a linear combination of elementary differentials of order $(\kappa_1+\ldots+\kappa_\tau+k)=\kappa$ where the coefficients are natural numbers. Done.

Corollary: it holds that

$$(1.34)\quad R_{i,\kappa-k}^{(k)} = \frac{(\kappa-k)!}{\kappa!} \sum_{\text{Ord } F=\kappa} \beta \psi_i^{(k)} [F]_o \quad .$$

Proof: Inserting (1.28) into (1.27) we obtain

$$\sum_{k=1}^{\kappa} \sum_{i=1}^{n} c_i^{(k)} \frac{\kappa!}{(\kappa-k)!} R_{i,\kappa-k}^{(k)} = \sum_{\text{ord } F=\kappa} \beta \sum_{k=1}^{\kappa} \sum_{i=1}^{n} c_i^{(k)} \psi_i^{(k)} [F]_o =$$

$$= \sum_{k=1}^{\kappa} \sum_{i=1}^{n} c_i^{(k)} \sum_{\text{ord } F=\kappa} \beta \psi_i^{(k)} [F]_o \quad ;$$

Hence follows (1.34).

This theorem suggests the following definition of so-called " elementary weights" (Butcher / 1 / p.194):

Definition: To each elementary differential $F=\{F_1 \ldots F_s\}$ corresponds an elementary weight $\phi=[\phi_1 \ldots \phi_i]$, where ϕ is the corresponding coefficient of F in (1.27).

Remarks:

1) In addition equations (1.1o) and (1.27) adjoin to each elementary differential F a number α and a number β. The correspondence of α, β, ϕ to F shall be expressed by similiar indices.
2) In contrast to F the coefficients α, β, and ϕ do not depend on the function f of the differential equation (1.1).

The next task is now the computation of ϕ and β:
for this we insert (1.34) into (1.32) and use the formula

$$\binom{\varkappa_i}{\sigma_i}\frac{(\varkappa_i-\sigma_i)!}{\varkappa_i!} = \frac{1}{\sigma_i!} \quad ;$$

$$R^{(k)}_{i,\varkappa-k} =$$

$$\sum_{\tau=1}^{\varkappa-k} \sum_{\substack{\varkappa_1+\ldots+\varkappa_\tau=\varkappa-k \\ \varkappa_1\geq 1}} \sum_{\text{ord } F_1=\varkappa_1} \cdots \sum_{\text{ord } F_\tau=\varkappa_\tau} \frac{(\varkappa-\tau)!}{\tau!}\frac{\beta_1}{\varkappa_1!}\cdots\frac{\beta_\tau}{\varkappa_\tau!}\cdot$$

$$\cdot\left(\sum_{\sigma_1=1}^{\varkappa_1}\frac{1}{\sigma_1!}\sum_{j_1=1}^{n} b_{ij_1}^{(\sigma_1)}\Psi_{1,j_1}^{(\sigma_1)}\right)\cdots\left(\sum_{\sigma_\tau=1}^{\varkappa_\tau}\frac{1}{\sigma_\tau!}\sum_{j_\tau=1}^{n} b_{ij_\tau}^{(\sigma_\tau)}\Psi_{\tau,j_\tau}^{(\sigma_\tau)}\right)\cdot\left[F_1\ldots F\frac{d^\tau}{\tau_{dy}{}^\tau}D^k y\right]_c =$$

$$\sum_{\tau=1}^{\varkappa-k} \sum_{\substack{\varkappa_1+\ldots+\varkappa_\tau=\varkappa-k \\ \varkappa_1\geq 1}} \sum_{\text{ord } F_1=\varkappa_1} \cdots \sum_{\text{ord } F_\tau=\varkappa_\tau} \frac{(\varkappa-k)!}{\tau!}\left(\prod_{l=1}^{\tau}\frac{\beta_l}{\varkappa_l!}\Psi_{l,i}^{(o)}\right)\cdot$$

(1.35)

$$\cdot\left[F_1\ldots F_\tau\frac{d^\tau}{dy^\tau}D^k y\right]_o$$

with

$$\Psi_{l,i}^{(o)} := \sum_{\sigma_l=1}^{\varkappa_l}\frac{1}{\sigma_l!}\sum_{j=1}^{n} b_{ij}^{(\sigma_l)}\Psi_{l,j}^{(\sigma_l)} \quad (l=1,\ldots,\tau)\ . \tag{1.36}$$

The insertion of (1.35) in (1.27) gives now with $\varkappa=r$:

$$\sum_{\text{ord } F=r}\beta\phi[F]_o =$$

$$\sum_{k=1}^{r}\sum_{i=1}^{n} o_i^{(k)} \sum_{\tau=1}^{r-k} \sum_{\substack{r_1+\ldots+r_\tau=r-k \\ r_l\geq 1}} \sum_{\text{ord } F_1=r_1} \cdots \sum_{\text{ord } F_\tau=r_\tau} \frac{r!}{\tau!}\left(\prod_{l=1}^{\tau}\frac{\beta_l}{\varkappa_l!}\Psi_{l,i}^{(o)}\right)\cdot$$

(1.37)

$$\cdot\left[F_1\ldots F_\tau\frac{d^\tau}{dy^\tau}D^k y\right]_o \quad .$$

Comparing the right and left side of this identity, we now obtain recursion formulas for ϕ and β: for, on the left side β and ϕ correspond to F, which is of order r; on the right side β_l and $\Psi_{l,i}^{(o)}$ correspond to F_l, whose order is smaller or equal to r-1.

Conditions for the Parameters

For m=1 in (1.15,16), hence for the classical Runge-Kutta-method , the comparison in (1.37) is easily done:
because of (1.17) we have

$$\psi^{(o)}_{1,i} = \sum_{j=1}^{n} b^{(1)}_{ij}\psi^{(1)}_{1,j} \quad \text{and} \quad c^{(k)}_i = o \text{ for } k=2,3,\dots .$$

Hence (1.37) becomes:

$$\sum_{\text{ord } F=r} \beta\phi[F]_o =$$

$$\sum_{i=1}^{n} c^{(1)}_i \sum_{\kappa=1}^{r-1} \sum_{\substack{r_1+\dots+r_\kappa=r-1 \\ r_1 \geq 1}} \sum_{\text{ord } F_1=r_1} \cdots \sum_{\text{ord } F_\kappa=r_\kappa} \frac{r!}{\kappa!} \left(\prod_{l=1}^{\sigma} \frac{\beta_l}{r_l!} \sum_{j=1}^{n} b^{(1)}_{ij}\psi^{(1)}_{l,j} \right) \cdot$$

$$\cdot \left[\{ F_1 \dots F_\kappa \} \right]_o$$

If we choose in the left hand side a fixed elementary differential, say $F = \{ F_1^{\mu_1} \dots F_\sigma^{\mu_\sigma} \}$ with $\sum_{i=1}^{\sigma} \mu_i = s$, this appears in the right hand side with the following coefficient:

$$\beta\phi[F]_o = \sum_{i=1}^{n} c^{(1)}_i \cdot \frac{r!}{s!} \left(\prod_{l=1}^{\sigma} \left(\frac{\beta_l}{r_l!} \right)^{\mu_l} \left(\sum_{j=1}^{n} b^{(1)}_{ij}\psi^{(1)}_{l,j} \right)^{\mu_l} \right) \cdot \frac{s!}{\mu_1! \dots \mu_\sigma!} [F]_o =$$

$$r! \prod_{l=1}^{\sigma} \frac{1}{\mu_l!} \left(\frac{\beta_l}{r_l!} \right)^{\mu_l} \cdot \sum_{i=1}^{n} c^{(1)}_i \prod_{l=1}^{\sigma} \left(\sum_{j=1}^{n} b^{(1)}_{ij}\psi^{(1)}_{l,j} \right)^{\mu_l} \cdot [F]_o \ .$$

The factor $\frac{s!}{\mu_1! \dots \mu_\sigma!}$ is due to the permutation which are to bear in mind in the sum $\sum_{r_1+\dots+r_s=r-1}$.

Comparison now gives

$$\beta = r! \prod_{l=1}^{\sigma} \frac{1}{\mu_l!} \left(\frac{\beta_l}{r_l!} \right)^{\mu_l} \qquad (1.38a)$$ [1]

$$\phi = \sum_{i=1}^{n} c^{(1)}_i \prod_{l=1}^{\sigma} \left(\sum_{j=1}^{n} b^{(1)}_{ij}\psi^{(1)}_{l,j} \right)^{\mu_l} \ . \qquad (1.38b)$$

These are exactly the formulas of Butcher / 1/, p. 194-196.

---

1) The coefficients β defined here differ from those of Butcher. His coefficients are obtained by puttin $\beta' = \beta/r$.

For m>1 this comparison is more complicated. First we need the following two Lemmas:

Lemma 1: Let u_i be functions of $y_1,\dots,y_n$ and sufficiently often differentiable; then with $\varkappa_i = \varkappa_{1i}+\dots+\varkappa_{si}$ it holds that

$$\frac{\partial^s}{\partial y_{j_1}\dots\partial y_{j_s}}(u_1\cdot u_2\dots u_t) = \tag{1.39}$$

$$\sum_{\substack{\varkappa_{11}+\dots+\varkappa_{1t}=1\\ \varkappa_{1i}\geq 0}}\cdots\sum_{\substack{\varkappa_{s1}+\dots+\varkappa_{st}=1\\ \varkappa_{si}\geq 0}} \frac{\partial^{\varkappa_1}u_1}{\partial y_{j_1}^{\varkappa_{11}}\dots\partial y_{j_s}^{\varkappa_{s1}}}\cdots\frac{\partial^{\varkappa_t}u_t}{\partial y_{j_1}^{\varkappa_{1t}}\dots\partial y_{j_s}^{\varkappa_{st}}} .$$

Proof (by induction on s):

s=1: $$\frac{\partial}{\partial y_{j_1}}(u_1\dots u_t) = \sum_{\substack{\varkappa_{11}+\dots+\varkappa_{1t}=1\\ \varkappa_{1i}\geq 0}} \frac{\partial^{\varkappa_1}u_1}{\partial y_{j_1}^{\varkappa_{11}}}\cdots\frac{\partial^{\varkappa_t}u_t}{\partial y_{j_t}^{\varkappa_{1t}}} = \text{(since } \varkappa_{1i}=\varkappa_i\text{)}$$

$$\frac{\partial u_1}{\partial y_{j_1}}u_2\dots u_t + u_1\frac{\partial u_2}{\partial y_{j_2}}u_3\dots u_t + \dots + u_1\dots u_{t-1}\frac{\partial u_t}{\partial y_{j_t}} .$$

Induction from s-1 to s:

$$\frac{\partial^s}{\partial y_{j_1}\dots\partial y_{j_s}}(u_1\dots u_t) = \frac{\partial}{\partial y_{j_1}}\left(\frac{\partial^{s-1}}{\partial y_{j_2}\dots\partial y_{j_t}}(u_1\dots u_t)\right) =$$

(by induction hypothesis with $\varkappa_i' = \varkappa_{2i}+\dots+\varkappa_{si}$)

$$= \frac{\partial}{\partial y_{j_1}}\left(\sum_{\substack{\varkappa_{21}+\dots+\varkappa_{2t}=1\\ \varkappa_{2i}\geq 0}}\cdots\sum_{\substack{\varkappa_{s1}+\dots+\varkappa_{st}=1\\ \varkappa_{si}\geq 0}} \frac{\partial^{\varkappa_1'}u_1}{\partial y_{j_2}^{\varkappa_{21}}\dots\partial y_{j_s}^{\varkappa_{s1}}}\cdots\frac{\partial^{\varkappa_t'}u_t}{\partial y_{j_2}^{\varkappa_{2t}}\dots\partial y_{j_s}^{\varkappa_{st}}}\right) =$$

(by commutation of $\frac{\partial}{\partial y_{j_1}}$ with summation sign and by (1.39) with s=1)

$$\sum_{\substack{\varkappa_{11}+\dots+\varkappa_{1t}=1\\ \varkappa_{1i}\geq 0}}\cdots\sum_{\substack{\varkappa_{s1}+\dots+\varkappa_{st}=1\\ \varkappa_{si}\geq 0}} \frac{\partial^{\varkappa_1'+\varkappa_{11}}u_1}{\partial y_{j_1}^{\varkappa_{11}}\dots\partial y_{j_s}^{\varkappa_{s1}}}\cdots\frac{\partial^{\varkappa_t'+\varkappa_{1t}}u_t}{\partial y_{j_1}^{\varkappa_{1t}}\dots\partial y_{j_s}^{\varkappa_{st}}} .$$

Putting now $\varkappa_i = \varkappa_i'+\varkappa_{1i} = \varkappa_{1i}+\dots+\varkappa_{si}$ we obtain (1.39).

Done.

For Lemma 2 we need a further symbol, which is now defined:

Defintion: Let $F=\{F_1^{\varkappa_1}\dots F_\pi^{\varkappa_\pi}\}=\{F_1\dots F_p\}$ with order r and $p=\varkappa_1+\dots+\varkappa_\pi$, and $\hat{F}=\{\hat{F}_1\dots\hat{F}_t\}$ with order $\hat{r}$ [1] be elementary differentials. Then we define:

(1.4oa) $$f^*(\varkappa_{ik};F_1,\dots,F_\pi)=\{1\}^*(\varkappa_{ik};F_1,\dots,F_\pi):=\{F_1^{\varkappa_{11}}\dots F_\pi^{\varkappa_{\pi 1}}\}$$

(1.4ob) $$\hat{F}^*(\varkappa_{ik};F_1,\dots,F_\pi)=\{\hat{F}_1\dots\hat{F}_t\}^*(\varkappa_{ik};F_1,\dots,F_\pi):=$$
$$\{F_1^{\varkappa_{11}^{(o)}}\dots F_\pi^{\varkappa_{\pi 1}^{(o)}}\hat{F}_1^*(\varkappa_{ik}^{(1)};F_1,\dots F_\pi)\dots\hat{F}_t^*(\varkappa_{ik}^{(t)};F_1,\dots$$

where

1.) $\varkappa_{ik}^{(1)}$ non negative integer;

2.) set of indices for $\varkappa_{ik}^{(1)}$:

(1.4oc) $$i=1,\dots,\pi;\quad k=1,\dots,\hat{r}_l;\quad l=o,1,\dots,t$$

and for $\varkappa_{ik}$:

(1.4od) $$i=1,\dots,\pi;\quad k=1,\dots,\hat{r}$$

(1.4oe) 3.) $\varkappa_{ik}^{(1)}=\varkappa_{im}$ with $m=\sum_{j=o}^{l-1}\hat{r}_j+k \quad (\hat{r}_o:=1)$.

Convention: If misunderstandings are not possible, we shall write for $\hat{F}^*(\varkappa_{ik};F_1,\dots,F_\pi)$ shortly $\hat{F}^*(\varkappa_{ik})$.

Example:

$F=\{F_1^{\varkappa_1}\dots F_\pi^{\varkappa_\pi}\}$ and $\hat{F}=\{\{f\}f\}$: hence $\hat{F}_1=\{f\}$, $\hat{F}_2=f$;

$\hat{r}_1=2$, $\hat{r}_2=1$, $\hat{r}=4$;

$\hat{F}_2^*(\varkappa_{ik}^{(2)})=\{F_1^{\varkappa_{11}^{(2)}}\dots F_\pi^{\varkappa_{\pi 1}^{(2)}}\}$ (by (1.4oa)) ,

$\hat{F}_1^*(\varkappa_{ik}^{(1)})=\{F_1^{\varkappa_{11}^{(1)}}\dots F_\pi^{\varkappa_{\pi 1}^{(1)}}\{F_1^{\varkappa_{12}^{(1)}}\dots F_\pi^{\varkappa_{\pi 2}^{(1)}}\}\}$

---

1) The quantities $\hat{r}$, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\Psi}_i$ correspond to $\hat{F}$; $\hat{r}_k$, $\hat{\alpha}_k$, $\hat{\beta}_k$, $\hat{\gamma}_k$, $\hat{\Psi}_{k,i}$ correspond to $\hat{F}_k$.

$$\hat{F}^*(\varkappa_{ik}) = \{ F_1^{\varkappa_{11}^{(0)}} \ldots F_\pi^{\varkappa_{\pi 1}^{(0)}} \{ F_1^{\varkappa_{11}^{(1)}} \ldots F_\pi^{\varkappa_{\pi 1}^{(1)}} \{F_1^{\varkappa_{12}^{(1)}} \ldots F_\pi^{\varkappa_{\pi 2}^{(1)}}\}\} \{ F_1^{\varkappa_{11}^{(2)}} \ldots F_\pi^{\varkappa_{\pi 1}^{(2)}} \}\} =$$

(with (1.4oe))

$$= \{ F_1^{\varkappa_{11}} \ldots F_\pi^{\varkappa_{\pi 1}} \{ F_1^{\varkappa_{12}} \ldots F_\pi^{\varkappa_{\pi 2}} \{ F_1^{\varkappa_{13}} \ldots F_\pi^{\varkappa_{\pi 3}}\}\} \{F_1^{\varkappa_{14}} \ldots F_\pi^{\varkappa_{\pi 4}}\}\} .$$

Lemma 2: Let $F = \{F_1^{\varkappa_1} \ldots F_\pi^{\varkappa_\pi}\} = \{ F_1 \ldots F_p \}$ and $\hat{F} = \{\hat{F}_1 \ldots \hat{F}_t\}$,

then it holds that

$$F_1^{\varkappa_1} \ldots F_\pi^{\varkappa_\pi} \frac{d^p}{dy^p} \hat{F} =$$

$$(1.41) \quad \sum_{\substack{\varkappa_{11}+\ldots+\varkappa_{1\hat{t}}=\varkappa_1 \\ \varkappa_{1k} \geq 0}} \ldots \sum_{\substack{\varkappa_{\pi 1}+\ldots+\varkappa_{\pi\hat{t}}=\varkappa_\pi \\ \varkappa_{\pi k} \geq 0}} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j!}{\varkappa_{j1}! \ldots \varkappa_{j\hat{t}}!} \right) \hat{F}^*(\varkappa_{ik}; F_1, \ldots F_\pi)$$

with $p = \varkappa_1 + \ldots + \varkappa_\pi$.

The elementary differentials $\hat{F}^*(\varkappa_{ik})$ have the order $r+\hat{r}-1$.

Proof (by induction on $\hat{r}$):

$\hat{r}=1$: Here $\hat{F} = f$ and $\varkappa_{ik} = \varkappa_i$ $(i=1,\ldots,\pi)$.

From (1.4oa) and (1.9) it follows that (1.41) is satisfied.

Next the following summation rule is valid:

$$1 \leq i_1 < i_2 < \ldots i_k < t = i_{k+1}$$

$$(1.42a) \quad \lambda_j = \sigma_{i_j+1} + \ldots + \sigma_{i_{j+1}} \qquad (j=1,\ldots,k)$$

then

$$\sum_{\sigma_1+\ldots+\sigma_t=\sigma} (\ldots) =$$

$$(1.42b) \quad \sum_{\sigma_1+\ldots+\sigma_{i_1}+\lambda_1+\ldots+\lambda_k=\sigma} \sum_{\sigma_{i_1+1}+\ldots+\sigma_{i_2}=\lambda_1} \ldots \sum_{\sigma_{i_k+1}+\ldots+\sigma_t=\lambda_k} (\ldots) .$$

For simplification we define the following sets of tupels:

$$(1.43) \quad K_k^{(i)} = \{ (\varkappa_{k1}^{(i)}, \ldots, \varkappa_{k\hat{r}_i}^{(i)}) \mid \varkappa_{kj}^{(i)} > 0 \ (j=1,\ldots,\hat{r}_i) \text{ and } \varkappa_{k1}^{(i)} + \ldots + \varkappa_{k\hat{r}_i}^{(i)} = \varkappa_k^{(i)} \} .$$

Induction from $\hat{r}-1$ to $\hat{r}$:

With (1.43) the induction hypothesis reads as follows:

$$F_1^{\varkappa_1^{(i)}} \dots F_\pi^{\varkappa_\pi^{(i)}} \frac{d^{\varkappa^{(i)}}}{dy^{\varkappa^{(i)}}} \hat{F}_i = \sum_{K_1^{(i)}} \dots \sum_{K_\pi^{(i)}} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j^{(i)}!}{\varkappa_{j1}^{(i)}! \dots \varkappa_{j\hat{r}_i}^{(i)}!} \right) \hat{F}_i^*(\varkappa_{ik}^{(i)}) \tag{1.44}$$

$$\text{with } \varkappa_1^{(i)} + \dots + \varkappa_\pi^{(i)} = \varkappa^{(i)} \quad , \quad (i=1,\dots,t) \quad .$$

Then

$$F_1^{\varkappa_1} \dots F_\pi^{\varkappa_\pi} \frac{d^p}{dy^p} \hat{F} = F_1 \dots F_p \frac{d^p}{dy^p} \left(\hat{F}_1 \dots \hat{F}_t \frac{d^t}{dy^t} f \right) \quad (1)$$

$$\sum_{\sigma_{10}+\dots+\sigma_{1t}=1} \dots \sum_{\sigma_{p0}+\dots+\sigma_{pt}=1} \left(F^{\sigma_{11}} \dots F_p^{\sigma_{p1}} \frac{d^{\varkappa^{(1)}}}{dy^{\varkappa^{(1)}}} \hat{F}_1 \right) \dots \left(F_1^{\sigma_{1t}} \dots F_p^{\sigma_{pt}} \frac{d^{\varkappa^{(t)}}}{dy^{\varkappa^{(t)}}} \hat{F}_t \right) \cdot$$

$$\cdot \left(F_1^{\sigma_{10}} \dots F_p^{\sigma_{p0}} \frac{d^{\varkappa^{(0)}+t}}{dy^{\varkappa^{(0)}+t}} f \right) \quad (2)$$

$$\sum_{\varkappa_1^{(0)}+\dots+\varkappa_1^{(t)}=\varkappa_1} \dots \sum_{\varkappa_\pi^{(0)}+\dots+\varkappa_\pi^{(t)}=\varkappa_\pi} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j!}{\varkappa_j^{(0)}! \dots \varkappa_j^{(t)}!} \right) \left(F_1^{\varkappa_1^{(1)}} \dots F_\pi^{\varkappa_\pi^{(1)}} \frac{d^{\varkappa^{(1)}}}{dy^{\varkappa^{(1)}}} \hat{F}_1 \right) \cdot$$

$$\dots \left(F_1^{\varkappa_1^{(t)}} \dots F_\pi^{\varkappa_\pi^{(t)}} \frac{d^{\varkappa^{(t)}}}{dy^{\varkappa^{(t)}}} \hat{F}_t \right) \cdot \left(F_1^{\varkappa_1^{(0)}} \dots F_\pi^{\varkappa_\pi^{(0)}} \frac{d^{\varkappa^{(0)}+t}}{dy^{\varkappa^{(0)}+t}} f \right) \quad (1.44)$$

$$\sum_{\varkappa_1^{(0)}+\dots+\varkappa_1^{(t)}=\varkappa_1} \dots \sum_{\varkappa_\pi^{(0)}+\dots+\varkappa_\pi^{(t)}=\varkappa_\pi} \sum_{K_1^{(1)}} \dots \sum_{K_\pi^{(1)}} \dots \sum_{K_1^{(t)}} \dots \sum_{K_\pi^{(t)}} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j!}{\varkappa_j^{(0)}! \dots \varkappa_j^{(t)}!} \cdot \right.$$

$$\left. \cdot \frac{\varkappa_j^{(1)}!}{\varkappa_{j1}^{(1)}! \dots \varkappa_{j\hat{r}_1}^{(1)}!} \cdot \dots \cdot \frac{\varkappa_j^{(t)}}{\varkappa_{j1}^{(t)}! \dots \varkappa_{j\hat{r}_t}^{(t)}!} \right) \cdot \left(\hat{F}_1^*(\varkappa_{1k}^{(1)}) \dots \hat{F}_t^*(\varkappa_{1k}^{(t)}) F_1^{\varkappa_1^{(0)}} \dots F_\pi^{\varkappa_\pi^{(0)}} \frac{d^{\varkappa^{(0)}+t}}{dy^{\varkappa^{(0)}+t}} f \right) =$$

(by rearranging the sum signs and using (1.9))

$$\sum_{\varkappa_1^{(0)}+\dots+\varkappa_1^{(t)}=\varkappa_1} \sum_{K_1^{(1)}} \dots \sum_{K_1^{(t)}} \dots \sum_{\varkappa_\pi^{(0)}+\dots+\varkappa_\pi^{(t)}=\varkappa_\pi} \sum_{K_\pi^{(1)}} \dots \sum_{K_\pi^{(t)}} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j!}{\varkappa_j^{(0)}! \varkappa_{j1}^{(1)}! \dots \varkappa_{j\hat{r}_t}^{(t)}!} \right) \cdot$$

$$\cdot \left\{ F_1^{\varkappa_1^{(0)}} \dots F_\pi^{\varkappa_\pi^{(0)}} \hat{F}_1^*(\varkappa_{ik}^{(1)}) \dots \hat{F}_t^*(\varkappa_{ik}^{(t)}) \right\} =$$

(using (1.42) and (1.40b))

$$\sum_{\varkappa_1^{(0)}+\varkappa_{11}^{(1)}+\dots+\varkappa_{1r_1}^{(t)}=\varkappa_1} \dots \sum_{\varkappa_\pi^{(0)}+\varkappa_{\pi 1}^{(1)}+\dots+\varkappa_{\pi r_\pi}^{(t)}=\varkappa_\pi} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j!}{\varkappa_j^{(0)}! \varkappa_{j1}^{(1)}! \dots \varkappa_{j\hat{r}_t}^{(t)}!} \right) \hat{F}^*(\varkappa_{ik}) =$$

(using (1.40c) and $\varkappa_1^{(0)} = \varkappa_{11}^{(0)}$)

$$\sum_{\varkappa_{11}+\dots+\varkappa_{1r}=\varkappa_1} \dots \sum_{\varkappa_{\pi 1}+\dots+\varkappa_{\pi r}=\varkappa_\pi} \left(\prod_{j=1}^{\pi} \frac{\varkappa_j!}{\varkappa_{j1}! \dots \varkappa_{j\hat{r}}!} \right) \hat{F}^*(\varkappa_{ik}) \quad .$$

(1) : This step is verified by transscribing into the sum-notation (cf. (1.7,9)) then by using (1.39) and finally by transscribing back with (1.7,9) .

(2) : This step is verified by collecting together in $F_1^{\sigma_{1i}}\dots F_p^{\sigma_{pi}}$ all elementary differentials which occur multiple;so we obtain $F_1^{\varkappa_1^{(i)}}\dots F_\pi^{\varkappa_\pi^{(i)}}$:

let $F_{k_1}=\dots=F_{k_{\varkappa_1}} = F_1$, where $F_{k_i}\in\{F_1,\dots,F_p\}$;

Then

$$\varkappa_1^{(i)}=\sum_{l=1}^{\varkappa_1}\sigma_{k_l i} \qquad (i=o,\dots,t) \qquad \text{and}$$

$$\varkappa_1=\sum_{j=o}^{t}\varkappa_1^{(j)} \quad .$$

Finally we comprise the sums with $\sigma_{k_l i}$ and obtain

$$\sum_{\substack{\sigma_{k_1 o}+\dots+\sigma_{k_1 t}=1\\ \sigma_{k_1 i}\geq o}}\dots\sum_{\substack{\sigma_{k_{\varkappa_1} o}+\dots+\sigma_{k_{\varkappa_1} t}=1\\ \sigma_{k_{\varkappa_1} i}\geq o}}(\dots)\ \text{(using (1.45a,b))}$$

$$=\sum_{\varkappa_1^{(o)}+\dots+\varkappa_1^{(t)}=\varkappa_1}\frac{\varkappa_1!}{\varkappa_1^{(o)}!\dots\varkappa_1^{(t)}!}(\dots) \quad .$$

Analoguously the exponent of $F_2,\dots F_\pi$ and the corresponding sums are comprised.

Next we prove by induction that

$\hat{F}^*(\varkappa_{ik})$ has the order $\sum_{j=1}^{\pi}\varkappa_j r_j+\hat{r}$, (where $\varkappa_j=\sum_{k=1}^{r}\varkappa_{jk}$):

induction hypothesis: $\hat{F}_i^*(\varkappa_{jk}^{(i)})$ has the order $\sum_{j=1}^{\pi}\varkappa_j^{(i)}r_j+\hat{r}_i$ (i=1,...,t);

then

$\hat{F}^*(\varkappa_{ik})=\{F_1^{\varkappa_{11}^{(o)}}\dots F_\pi^{\varkappa_{\pi 1}^{(o)}}\hat{F}_1^*(\varkappa_{ik}^{(1)})\dots\hat{F}_t^*(\varkappa_{ik}^{(t)})\}$ has the order (by (1.6b)):

$$\sum_{i=o}^{t}\sum_{j=1}^{\pi}\varkappa_j^{(i)}r_j+\hat{r}_1+\dots+\hat{r}_t+1=\sum_{j=1}^{\pi}(\sum_{i=o}^{t}\varkappa_j^{(i)})r_j+\hat{r}\ \overset{(1.45b)}{=}\sum_{j=1}^{\pi}\varkappa_j r_j+\hat{r} \quad .$$

Finally $F=\{F_1^{\varkappa_1}\dots F_\pi^{\varkappa_\pi}\}$ has the order $\sum_{j=1}^{\pi}\varkappa_j r_j+1=r$, and hence

$\hat{F}^{*}(x_{ik})$ has the order $r+\hat{r}-1$.

With this the proof of Lemma 2 is completed. Done.

Corollary: $F_1 \ldots F_\varkappa \frac{d^\varkappa}{dy^\varkappa} D^k y$ is a linear combination of elementary differentials of order $r_1+\ldots r_\varkappa+k$. This has been used in connection with (1.33).

We are now ready to prove:

Theorem: The elementary weight $\phi = [\phi_1 \ldots \phi_s]$ corresponding to $F = \{F_1 \ldots F_s\}$ is determined recursively by the following formulas:

ϕ, corresponding to f, is given by

$$\phi = \sum_{i=1}^{n} c_i^{(1)} \quad ; \tag{1.46a}$$

ϕ, corrsponding to F, is given by:

$$\phi = [\phi_1 \ldots \phi_s] = \sum_{k=1}^{m} \sum_{i=1}^{n} c_i^{(k)} \sum_{\substack{\lambda_1+\ldots+\lambda_t=k-1 \\ \lambda_l \geq 0}} \frac{(k-1)!}{\lambda_1! \ldots \lambda_s!} \psi_{1,i}^{(\lambda_1)} \ldots \psi_{s,i}^{(\lambda_s)} \tag{1.46b}$$

where $\psi_{l,i}^{(\lambda_l)}$ ($\lambda_l \geq 1$) correspond to ϕ_l (cf. (1.28)) and

$$\psi_{l,i}^{(0)} = \sum_{\sigma=1}^{m} \frac{1}{\sigma!} \sum_{j=1}^{n} b_{ij}^{(\sigma)} \psi_{l,j}^{(\sigma)} \qquad l=1,\ldots,t \quad . \tag{1.36}$$

The coefficient β corresponding to $F = \{F_1^{\mu_1} \ldots F_\sigma^{\mu_\sigma}\}$ is given by:

$$\beta = r! \prod_{j=1}^{\sigma} \frac{1}{\mu_j!} \left(\frac{\beta_j}{r_j!}\right)^{\mu_j} \quad \text{where } \beta_j \text{ corresponds to } F_j. \tag{1.47}$$

Proof:

1.) (1.46a) follows directly from (1.27), since $\varkappa=1$ and hence $k=1$ and $R_{i,o}^{(1)} \overset{(1.20a)}{=} [f]_o$: so $\phi = \sum_{i=1}^{n} c_i^{(1)}$.

2.) Let

$F = \{F_1^{\mu_1} \ldots F_\sigma^{\mu_\sigma}\}$, $F_i = \{F_{i1}^{\mu_{i1}} \ldots F_{i\sigma_i}^{\mu_{i\sigma_i}}\}$, $F_{ik} = \{\ldots\}$ and so on, until f is reached.

Out of all F_{ik}, F_{ikl}, ...,f we now collect these elementary differentials, which differ from $F_1,\dots,F_\sigma$ and denote them by $F_{\sigma+1},\dots,F_\pi$. Then each of the elementary differentials of which F is composed can be written in the form $\{F_1^{\varkappa_1}\dots F_\pi^{\varkappa_\pi}\}$ where $\varkappa_i \geq 0$.

Example:

$F= \{\{f\}\{f\{f^2\}\}\}$:

$F_1= \{f\}$, $F_2= \{f\{f^2\}\}$, $F_3=\{f^2\}$, $F_4= f$;

then for example:

$F= \{F_1F_2F_3^0F_4^0\}$, $F_2= \{F_1^0F_2^0F_3F_4\}$.

3.) Let $M(\hat{F};F)$ be the following set of tupels:

$$(1.48)\quad M(\hat{F};F) := \{(\varkappa_{11},\dots,\varkappa_{\pi\hat{r}}) \mid \hat{F}^*(\varkappa_{ik};F_1,\dots,F_\pi)= F\} .$$

Choosing one tupel out of $M(\hat{F};F)$ and putting

$$(1.49a)\quad \varkappa_i= \varkappa_{i1}+\dots+\varkappa_{i\hat{r}} \;;(i=1,\dots,\pi) \quad \text{and} \quad p= \varkappa_1+\dots+\varkappa_\pi ,$$

we obtain from Lemma 2

$$(1.49b)\quad F_1^{\varkappa_1}\dots F_\pi^{\varkappa_\pi}\frac{d^p}{dy^p}\hat{F}= \left(\prod_{j=1}^{\pi}\frac{\varkappa_j!}{\varkappa_{j1}!\dots\varkappa_{j\hat{r}}!}\right)\cdot F + \dots$$

4.) From (1.1o) we have

$$(1.5o)\quad F_1\dots F_\lambda\frac{d^\lambda}{dy^\lambda}D^k y = \sum_{\text{ord } \hat{F}=k}\hat{\alpha}\left(F_1\dots F_\lambda\frac{d^\lambda}{dy^\lambda}\hat{F}\right)$$

5.) Now the comparison of terms in (1.37) is possible:

In the left side of (1.37) we choose an arbitrary elementary differential $F= \{F_1^{\mu_1}\dots F_\sigma^{\mu_\sigma}\}$ of order r.

In the right hand side of (1.37) we now are interested only in those terms, which contains F:

for this, we insert (1.5o) into (1.37) and let the summation run only on the tupels of $M(\hat{F};F)$; so we obtain by bearing in mind the number of permutations of the sum $\sum_{r_1+\dots+r_\varkappa=r-k}$ and by (1.49a,b)

$$\beta\phi[F]_o = \sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{\text{ord } \hat{F}=k} \hat{\alpha} \sum_{M(\hat{F};F)} \frac{r!}{p!}\left(\prod_{j=1}^{\pi}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_j}\right)\cdot\left(\prod_{j=1}^{\pi}\frac{\varkappa_j!}{\varkappa_{j1}!\dots\varkappa_{jk}!}\right)\cdot$$

$$\cdot\frac{p!}{\varkappa_1!\dots\varkappa_\pi!}\cdot[F]_o =$$

$$r!\sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{\text{ord } \hat{F}=k} \hat{\alpha} \sum_{M(\hat{F};F)}\left(\prod_{j=1}^{\pi}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_j}\frac{1}{\varkappa_{j1}!\dots\varkappa_{jk}!}\right)\cdot[F]_o \qquad (1.49a)$$

$$r!\sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{\text{ord } \hat{F}=k} \hat{\alpha} \sum_{M(\hat{F};F)} \prod_{l=1}^{k}\left(\prod_{j=1}^{\pi}\frac{1}{\varkappa_{jl}!}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_{jl}}\right)\cdot[F]_o \quad .$$

Putting

$$(1.51)\qquad \Lambda_i^{(k)} = \sum_{\text{ord } \hat{F}=k} \hat{\alpha} \sum_{M(\hat{F};F)} \prod_{l=1}^{k}\left(\prod_{j=1}^{\pi}\frac{1}{\varkappa_{jl}!}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_{jl}}\right)$$

we obtain

$$\beta\phi[F]_o = r!\sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)}\Lambda_i^{(k)}\cdot[F]_o \quad ;$$

using (1.28,29) it follows that

$$(1.52)\qquad \Lambda_i^{(k)} = \frac{\beta}{r!}\,\psi_i^{(k)}$$

This, however, does not jet lead to formulas for $\psi_i^{(k)}$, since usually the set $M(\hat{F};F)$ and therefor $\Lambda_i^{(k)}$ are not known. We thus try to find a recursive determination:

6.) We assume that the elementary weights $\phi_1,\dots,\phi_\sigma$ which correspond to $F_1,\dots,F_\sigma$ of $F=\{F_1^{\mu_1}\dots F_\sigma^{\mu_\sigma}\}$ are known; then also the following sets are known:

$$(1.53)\qquad M^{(1)}(\hat{F}_1;F_{j_1}) = \{\,(\varkappa_{11}^{(1)}\dots\varkappa_{\pi\hat{r}_1}^{(1)}) \mid \hat{F}_1^*(\varkappa_{ik}^{(1)};F_1,\dots,F_\pi) = F_{j_1}\,\}$$
$$(l=1,\dots,t\ ,\ 1\le j_1\le\sigma)$$

and hence also

$$(1.54)\qquad \Lambda_{j_1,i}^{(k)} = \frac{\beta_{j_1}}{r_{j_1}!}\,\psi_{j_1,i}^{(k)} \quad .$$

Using (1.53) it is now possible to construct the set $M(\hat{F};F)$: because of (1.4ob) we have

$$\hat{F}^*(\varkappa_{ik}) = \{\,F_1^{\varkappa_{11}^{(o)}}\dots F_\pi^{\varkappa_{\pi 1}^{(o)}}\,\hat{F}_1^*(\varkappa_{ik}^{(1)})\dots\hat{F}_t^*(\varkappa_{ik}^{(t)})\} \quad .$$

Since we want $\hat{F}^*(\varkappa_{ik}) = \{F_1^{\mu_1}\dots F_\sigma^{\mu_\sigma}\} = F$, it follows that:

$\hat{F}_1^*(\varkappa_{ik}^{(1)}) = F_{j_1}$, ... , $\hat{F}_t^*(\varkappa_{ik}^{(t)}) = F_{j_t}$ where $1 \leqslant j_1,\dots,j_t \leqslant \sigma$.

Since the tupels $(\dots,\varkappa_{jk}^{(i)},\dots)$ are known from (1.53), the still unknown numbers $\varkappa_{i1}^{(o)}$ are now obtained from the comparison

$$(1.55)\quad \{F_1^{\varkappa_{11}^{(o)}}\dots F_\pi^{\varkappa_{\pi 1}^{(o)}} F_{j_1}\dots F_{j_t}\} = \{F_1^{\mu_1}\dots F_\sigma^{\mu_\sigma}\}$$

as follows:

Let $\varrho_i \geqslant o$ $(i=1,2,\dots,\sigma)$ be the frequency of i occuring in $(j_1,\dots,j_t)$, then (1.55) gives:

$$(1.56)\quad \varkappa_{i1}^{(o)} = \mu_i - \varrho_i \ .$$

Hence it follows that only those $(j_1,\dots,j_t)$ can occur in (1.55) for which $\varkappa_{i1}^{(o)} \geqslant o$ $(i=1,\dots,\pi)$

Thus the set $M(\hat{F};F)$ can now be written

$$(1.57)\quad M(\hat{F};F) = \bigcup_{j_1,\dots,j_t=1}^{\sigma} \Big\{ (\varkappa_{11}^{(o)},\dots,\varkappa_{\pi 1}^{(o)},\dots \varkappa_{ik}^{(1)}\dots,\dots,\dots \varkappa_{ik}^{(t)}\dots) \Big| \varkappa_{i1}^{(o)} = \mu_i - \varrho_i \geqslant o \ (i=1,\dots,\pi),\ (\dots \varkappa_{ik}^{(l)} \dots) \in M^{(l)}(\hat{F}_l;F_{j_l}) \ (l=1,\dots,t)$$

/.) From (1.13) and (1.12) it follows that

$$F_1\dots F_s \frac{d^s}{dy^s} D^k y =$$

$$(1.58)\quad \sum_{t=1}^{k-1} \sum_{\substack{\lambda_1+\dots\lambda_t=k-1 \\ \lambda_i \geqslant 1}} \frac{(k-1)!}{t!} \cdot \frac{1}{\lambda_1!\dots\lambda_t!} \sum_{\text{ord } \hat{F}_1=\lambda_1} \dots \sum_{\text{ord } \hat{F}_t=\lambda_t} \hat{\alpha}_1\dots\hat{\alpha}_t \cdot \cdot F_1\dots F_s \frac{d^s}{dy^s}(\hat{F}_1\dots\hat{F}_t)$$

We insert (1.58) into (1.37) and let the summation run only over the tupels of $M(\hat{F};F)$;doing this we use again (1.49) and bear in mind, as in 5.) the number of permutations. Similar computations as in 5.) now lead to

$$\beta\phi[F]_o =$$

$$\sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{t=1}^{k-1} \sum_{\substack{\lambda_1+\ldots+\lambda_t=k-1\\ \lambda_1\geq 1}} \frac{(k-1)!}{t!}\cdot\frac{r!}{\lambda_1!\ldots\lambda_t!} \sum_{\text{ord }\hat{F}_1=\lambda_1}\ldots\sum_{\text{ord}\hat{F}_t=\lambda_t}\hat{\alpha}_1\ldots\hat{\alpha}_t\cdot$$

$$\sum_{j_1,\ldots,j_t=1}^{\sigma}\left(\prod_{j=1}^{\sigma}\frac{1}{\varkappa_{j1}^{(o)}}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_{j1}^{(o)}}\right)\cdot\sum_{M^{(1)}(\hat{F}_1;F_{j_1})}\ldots\sum_{M^{(t)}(\hat{F}_t;F_{j_t})}\cdot$$
$$\mu_1-\varrho_1\geq o$$

$$\cdot\prod_{\tau=1}^{t}\cdot\left(\prod_{l=1}^{\lambda_\tau}\prod_{j=1}^{\pi}\frac{1}{\varkappa_{jl}^{(\tau)}}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_{jl}^{(\tau)}}\right)\cdot[F]_o =$$

(by rearranging)

$$\sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{t=1}^{k-1} \sum_{\substack{\lambda_1+\ldots+\lambda_t=k-1\\ \lambda_1\geq 1}} \frac{(k-1)!}{t!}\cdot\frac{r!}{\lambda_1!\ldots\lambda_t!} \sum_{\substack{j_1,\ldots,j_t=1\\ \mu_i-\varrho_i\geq o}}\cdot$$

$$\cdot\left(\prod_{j=1}^{\sigma}\frac{1}{\varkappa_{j1}^{(o)}}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_{j1}^{(o)}}\right)\cdot\left[\prod_{\tau=1}^{t}\underbrace{\left(\sum_{\text{ord }\hat{F}_\tau=\lambda_\tau}\hat{\alpha}_\tau\sum_{M^{(\tau)}(\hat{F}_\tau;F_{j_\tau})}\prod_{l=1}^{\lambda_\tau}\prod_{j=1}^{\pi}\frac{1}{\varkappa_{jl}^{(\tau)}!}\left(\frac{\beta_j}{r_j!}\psi_{j,i}^{(o)}\right)^{\varkappa_{jl}^{(\tau)}}\right)}_{\Lambda_{j_\tau,i}^{(\lambda_\tau)}\ \text{(cf. (1.51))}}\right]$$

$$\cdot[F]_o \quad \text{(by using (1.54,56))}$$

$$=\left(r!\prod_{j=1}^{\sigma}\frac{1}{\mu_j!}\left(\frac{\beta_j}{r_j!}\right)^{\mu_j}\right)\cdot\left(\sum_{k=1}^{r}\sum_{i=1}^{n}c_i^{(k)}\sum_{t=1}^{k-1}\sum_{\substack{\lambda_1+\ldots+\lambda_t=k-1\\ \lambda_1\geq 1}}\frac{(k-1)!}{t!}\cdot\frac{1}{\lambda_1!\ldots\lambda_t!}\cdot\right.$$

$$\left.\sum_{j_1,\ldots,j_t=1}^{\sigma}\left(\prod_{l=1}^{\sigma}(\mu_l;-1;\varrho_l)\left(\psi_{l,i}^{(o)}\right)^{\mu_l-\varrho_l}\right)\psi_{j_1,i}^{(\lambda_1)}\ldots\psi_{j_t,i}^{(\lambda_t)}\right)\cdot[F]_o \quad ;$$

here we have used the following symbol (cf. Gröbner - Hofreiter: Integraltafel)

(1.59a) $\quad (a;-1;\lambda)= a(a-1)\ldots(a-\lambda+1) \quad (\lambda=1,2,\ldots)$

(1.59b) $\quad (a;-1;o)= 1$

The comparison of the coefficients in the above formulas gives

(1.47) $$\beta= r!\prod_{j=1}^{\sigma}\frac{1}{\mu_j!}\left(\frac{\beta_j}{r_j!}\right)^{\mu_j}$$

$$\phi = [\phi_1^{\mu_1}\dots\phi_\sigma^{\mu_\sigma}] = \sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{t=1}^{k-1} \sum_{\substack{\lambda_1+\dots+\lambda_t=k-1 \\ \lambda_1 \geq 1}} \frac{(k-1)!}{t!}\cdot\frac{1}{\lambda_1!\dots\lambda_t!}\cdot$$

(1.59)

$$\sum_{j_1,\dots,j_t=1}^{\sigma}\left(\prod_{l=1}^{\sigma}(\mu_l;-1;\varrho_l)\left(\Psi_{1,i}^{(o)}\right)^{\mu_l-\varrho_l}\right)\cdot\Psi_{j_1,i}^{(\lambda_1)}\dots\Psi_{j_t,i}^{(\lambda_t)}$$

These are the wanted recursion formulas for β and ϕ.

8.) Formula (1.59) can be simplified with the help of the following transscription:

(1.60a)

$$\phi = [\phi_1^{\mu_1}\phi_2^{\mu_2}\dots\phi_\sigma^{\mu_\sigma}]$$

$$\hat{\beta} = [\hat{\beta}_1\dots\hat{\beta}_{\pi_2-1}\hat{\beta}_{\pi_2}\dots\hat{\beta}_{\pi_3-1}\dots\hat{\beta}_{\pi_\sigma}\dots\hat{\beta}_s] = [\hat{\beta}_1\dots\hat{\beta}_s]$$

with

(1.60b) $$s = \sum_{i=1}^{\sigma}\mu_i \quad , \quad \pi_i = \sum_{k=1}^{i-1}\mu_k \quad (i=2,\dots,\sigma) \quad , \quad \pi_1 := 1 \quad .$$

Now formula (1.59) becomes

(1.46) $$[\hat{\beta}_1\dots\hat{\beta}_s] = \sum_{k=1}^{r}\sum_{i=1}^{n} c_i^{(k)} \sum_{\substack{\varkappa_1+\dots+\varkappa_s=k-1 \\ \varkappa_1 \geq 0}} \frac{(k-1)!}{\varkappa_1!\dots\varkappa_s!}\,\hat{\Psi}_{1,i}^{(\varkappa_1)}\dots\hat{\Psi}_{s,i}^{(\varkappa_s)} \quad .$$

Proof:

a) to prove this we need the follwing two summation rules:

First Rule:

Assume that out of $(\lambda_1,\dots,\lambda_t)$ τ numbers, say $\lambda_1,\dots,\lambda_\tau$ are distinct and that λ_i $(i=1,\dots,\tau)$ occur δ_i-times in $(\lambda_1,\dots,\lambda_t)$; if further in $(i_1,\dots,i_t)$ all i_k are distinct, it holds that

(1.62) $$\sum_{(\lambda_1,\dots,\lambda_t)} A_{i_1}^{(\lambda_1)}\dots A_{i_t}^{(\lambda_t)} = \frac{1}{\delta_1!\dots\delta_\tau!}\sum_{(i_1,\dots,i_t)} A_{i_1}^{(\lambda_1)}\dots A_{i_t}^{(\lambda_t)} \quad .$$

Here the symbol $\sum_{(\dots)}$ denotes summation over all permutations of $(\dots)$.

Second Rule:

(1.62) $$\sum_{j_1,\dots,j_t=1}^{\sigma}(\dots) = \sum_{1\leq j_1\leq\dots\leq j_t\leq\sigma}\ \sum_{(j_1,\dots,j_t)}(\dots)$$

b) In (1.59) we now fix k and t and choose a tupel $(\lambda_1,\dots,\lambda_t)$, which satisfied the same condition of a) . Then it follows from (1.59) bearing in mind the permutations of the sum $\sum_{\lambda_1+\dots+\lambda_t=k-1}$:

$$(1.63)\quad \sum_{i=1}^{n} c_i^{(k)} \frac{(k-1)!}{\lambda_1!\dots\lambda_t!} \cdot \frac{1}{\delta_1!\dots\delta_\tau!} \sum_{j_1,\dots,j_t=1}^{\sigma} \left(\prod_{l=1}^{\sigma} (\mu_l;-1;\varrho_l)\right) \left(\Psi_{1,i}^{(o)}\right)^{\mu_1-\varrho_1} \cdot \Psi_{j_1,i}^{(\lambda_1)} \dots \Psi_{j_t,i}^{(\lambda_t)} .$$

c) To show the equivalence of (1.46) and (1.59) it suffices to fix k and t and to sum up in (1.46) over the following tupels

$$(1.64)\quad (\varkappa_1,\dots,\varkappa_s) = (o,\dots,\varkappa_{i_1},\dots,o,\dots,\varkappa_{i_k},\dots,o,\dots,\varkappa_{i_t},\dots,o,\dots)$$
$$\text{with } \varkappa_{i_k} = \lambda_k \geq 1 \text{ and } 1 \leq i_1 < \dots < i_t \leq \sigma \quad .$$

The resulting expression has to equalize (1.63) :
With (1.64) it follows from (1.46) :

$$\sum_{i=1}^{n} c_i^{(k)} \sum_{1\leq i_1<i_2<\dots i_t\leq s} \sum_{(\lambda_1,\dots,\lambda_t)} \frac{(k-1)!}{\lambda_1!\dots\lambda_t!} \hat{\Psi}_{1,i}^{(o)} \dots \hat{\Psi}_{i_1,i}^{(\lambda_1)} \dots \hat{\Psi}_{1,i}^{(o)} \dots \hat{\Psi}_{i_t,i}^{(\lambda_t)} \dots \hat{\Psi}_{s,i}^{(o)} \quad (1.61)$$

$$\sum_{i=1}^{n} c_i^{(k)} \sum_{1\leq i_1<\dots<i_t\leq s} \sum_{(i_1,\dots,i_t)} \frac{(k-1)!}{\lambda_1!\dots\lambda_t!} \cdot \frac{1}{\delta_1!\dots\delta_\tau!} \hat{\Psi}_{1,i}^{(o)} \dots \hat{\Psi}_{i_1,i}^{(\lambda_1)} \dots \hat{\Psi}_{1,i}^{(o)} \dots \hat{\Psi}_{i_t,i}^{(\lambda_t)} \dots \hat{\Psi}_{s,i}^{(o)} \quad (1.65)$$
$$(\ 1\neq i_j \ j=1,\dots,t)$$

Next we turn over to the notation with $\Psi_{ik}^{(\cdot)}$ by putting

$i_k \dashrightarrow j_k$, $k=1,\dots,t$

$\hat{\Psi}_{k,i}^{(o)} \dashrightarrow \Psi_{l,i}^{(o)}$ where $\pi_l \leq k < \pi_{l+1}$ $(l=1,\dots,\sigma)$ (cf. (1.60a,b)) .

In $(i_1,\dots,i_t)$ all numbers are distinct, but not in $(j_1,\dots,j_t)$, since $1 \leq j_1,\dots,j_t \leq \sigma$. Again denote by ϱ_l $(l=1,\dots,\sigma)$ the frequency, with which l appears in $(j_1,\dots,j_t)$.
From (1.60a) we see:

if $\varrho_l > 0$, then exists a $\varkappa$ so that

$\pi_l \leqslant i_\varkappa < \ldots < i_{\varkappa+\varrho_l} < \pi_{l+1}$ $(l=1,\ldots,\sigma)$ and

$j_\varkappa = \ldots = j_{\varkappa+\varrho_l} = l$.

Thus, by the substitution $i_k \dashrightarrow j_k$ in (1.65)

$\sum_{1 \leqslant i_1 < \ldots < i_t \leqslant s} (\ldots)$ changes into $\sum_{1 \leqslant j_1 \leqslant \ldots \leqslant j_t \leqslant \sigma} \binom{\mu_1}{\varrho_1} \ldots \binom{\mu_\sigma}{\varrho_\sigma} (\ldots$

and

$\sum_{(i_1,\ldots,i_t)} (\ldots)$ changes into $\sum_{(j_1,\ldots,j_t)} \varrho_1! \ldots \varrho_\sigma! (\ldots)$.

Finally with $\binom{\mu_i}{\varrho_i} \varrho_i! = (\mu_i;-1;\varrho_i)$ (1.65) becomes

$$\sum_{i=1}^{n} c_i^{(k)} \frac{(k-1)!}{\lambda_1! \ldots \lambda_t!} \cdot \frac{1}{\delta_1! \ldots \delta_\tau!} \sum_{1 \leqslant j_1 \leqslant \ldots \leqslant j_t \leqslant \sigma} \sum_{(j_1,\ldots,j_t)} (\mu_1;-1;\varrho_1) \ldots$$

$$\ldots (\mu_\sigma;-1;\varrho_\sigma) \left(\psi_{1,i}^{(o)}\right)^{\mu_1-\varrho_1} \ldots \left(\psi_{\sigma,i}^{(o)}\right)^{\mu_\sigma-\varrho_\sigma} \cdot \psi_{j_1,i}^{(\lambda_1)} \ldots \psi_{j_t,i}^{(\lambda_t)} \quad .$$

But this, using (1.62), is equal to (1.63).

Thus, the proof of the theorem is completed. Done.

Remarks :

1) From (1.46) it can be seen, that the correspondence between $[\phi_1 \ldots \phi_s]$ and $\{F_1 \ldots F_s\}$ is one to one.
2) For m=1, hence k=1 and $\varkappa_1 = \ldots = \varkappa_s = o$ the formulas of Butcher (cf.(1.38)) are contained as special cases in (1.46).

Definition: A method for the integration of ordinary differential equations is said to have order p (or error-order p+1), if for its approximate solution $\hat{y}(x)$ holds that

(1.66) $$y(x) - \hat{y}(x) = O(h^{p+1})$$

for each solution y(x) of an arbitrary differential equation.

Theorem: A Runge-Kutta-process has order p if and only if the coefficients satisfy the conditions

(1.67) $\phi = \frac{1}{\gamma} = \frac{\alpha}{\beta}$ for all elemtabary differentials of order $r \leq p$.

The constants γ satisfy the recursion formula:

(1.68) $\gamma = r\gamma_1^{\mu_1} \ldots \gamma_\sigma^{\mu_\sigma}$ 1)

where γ corresponds to $\phi = [\phi_1^{\mu_1} \ldots \phi_\sigma^{\mu_\sigma}]$ and γ_i to ϕ_i .

Proof: The assertion follows from

$$y(x) = \sum_{k=0}^{\infty} \frac{h^k}{k!} \sum_{\text{ord } F=k} \alpha[F]_0 \qquad \text{(cf. (1.3) and (1.1o)) and}$$

$$\hat{y}(x) = \sum_{k=0}^{\infty} \frac{h^k}{k!} \sum_{\text{ord } F=k} \beta\phi[F]_0 \qquad \text{(cf. (1.22) and (1.27))}.$$

Thus the Taylor series coincide up to order p iff (1.67) is satisfied.

Formula (1.68) follows from (1.11) and (1.47) :

$$\gamma = \frac{\beta}{\alpha} = r \prod_{j=1}^{\sigma} \left(\frac{\beta_j}{\alpha_j}\right)^{\mu_j} = r\gamma_1^{\mu_1} \ldots \gamma_\sigma^{\mu_\sigma} .$$

Done.

Examples of Conditions

In many cases formula (1.46) for the elemtary weight can be simplified. For this we give some example, which shall be needed in the next sections:

We put

(1.7o) $$a_i = \sum_{j=1}^{n} b_{ij}^{(1)} \quad ;$$

this is motivated from transsribing (1.15) to non autonomous systems using:

---

1) The coefficients γ are also tabulated in Butcher / 1/,p.191-193.

$$(1.15') \quad g_i^{(k)} = (D^k y)\left(x_0+a_i h, y_0+h\sum_{j=1}^{n} b_{ij}^{(1)} g_j^{(1)} + \ldots + \frac{h^m}{m!}\sum_{j=1}^{n} b_{ij}^{(m)} g_j^{(m)}\right)$$

with

$$(1.4') \quad D = \frac{\partial}{\partial x} + \sum_j f_j \frac{\partial}{\partial y_j} \quad .$$

Examples:

$$(1.71) \quad \phi = \sum_{i=1}^{n} c_i^{(1)} = 1$$

$$(1.72) \quad [\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} (k;-1;\varkappa-1)\, a_i^{k-\varkappa+1} = \frac{1}{k+1}$$

$$(1.73) \quad [\phi_1\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1}\binom{\varkappa-1}{\sigma}(k;-1;\sigma)\, a_i^{k-\sigma}\psi_{1,i}^{(\varkappa-1-\sigma)} = \frac{1}{(k+r_1+1)\gamma_1}$$

$$(1.74) \quad [[\phi^l]\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} a_i^{k-\varkappa+1}\left(\sum_{\sigma=1}^{l}\frac{(l;-1;\sigma-1)(k;-1;\varkappa-1)}{\sigma!}\cdot\right.$$

$$\left.\cdot\sum_{j=1}^{n} b_{ij}^{(\sigma)} a_j^{l-\sigma+1} + \sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma}(k;-1;\sigma)(l;-1;\varkappa-\sigma-2)\, a_i^{l+1}\right) =$$

$$= \frac{1}{(k+l+2)(l+1)}$$

$$(1.75a) \quad [\phi_1\ldots\phi_s\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1}\binom{\varkappa-1}{\sigma}(k;-1;\sigma) a_i^{k-\sigma}\psi_i^{(\varkappa-\sigma)} = \frac{1}{(r+k)\gamma_1\ldots\gamma_s}$$

with

$$(1.75b) \quad \phi = [\phi_1\ldots\phi_s] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)}\psi_i^{(\varkappa)} \quad \text{and} \quad \gamma = r\gamma_1\ldots\gamma_s \quad .$$

Proof:

1.) (1.71) is already proofed (cf. (1.46a)).
The quantities ψ_i which correspond to ϕ are:

$$(1.76) \quad \psi_i^{(1)} = 1 \quad ; \quad \psi_i^{(k)} = 0 \,(k=2,3,\ldots) \quad ; \quad \psi_i^{(0)} = \sum_{j=1}^{n} b_{ij}^{(1)} = a_i \quad .$$

2.) Because of (1.76) for $[\phi^k]$ summation in (1.46) runs only over the following tupels und their permutations:

$(\underbrace{1,\dots,1}_{\varkappa-1},\underbrace{0,\dots,0}_{k-\varkappa-1})$.

$$[\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{(1,\dots,1,0,\dots,0)} (\varkappa-1)!\, a_i^{k-\varkappa+1} =$$

$$\sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \binom{k}{\varkappa-1} (\varkappa-1)!\, a_i^{k-\varkappa+1} = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} (k;-1;\varkappa-1) a_i^{k-\varkappa+1} .$$

3.) Because of (1.76) for $[\phi_1\phi^k]$ the summation in (1,46) is only over the tupels

$(\lambda_1,\dots,\lambda_s) = (\sigma,1,\dots,1,0,\dots,0)$ with $s=k+1$ and $0 \leq \sigma \leq \varkappa-1$

and their permutations:

hence

$$[\phi_1\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1} \underbrace{\binom{k}{\varkappa-\sigma-1}\frac{(\varkappa-1)!}{\sigma!}}_{\binom{\varkappa-1}{\sigma}(k;-1;\varkappa-\sigma-1)} \psi_{1,i}^{(\sigma)} a_i^{k-(\varkappa-\sigma-1)} \quad (\sigma \to \varkappa-\sigma-1)$$

$$= \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1} \binom{\varkappa-1}{\sigma} (k;-1;\sigma) a_i^{k-\sigma} \psi_{1,i}^{(\varkappa-\sigma-1)} .$$

4.) In (1.73) we put $\phi_1 = [\phi^1]$. Thus with (1.72):

$$\psi_{1,i}^{(\varkappa)} = (1;-1;\varkappa-1) a_i^{1-\varkappa+1} \quad (\varkappa=1,2,\dots)$$

$$\psi_{1,i}^{(0)} = \sum_{\sigma=1}^{m} \frac{1}{\sigma!} \sum_{j=1}^{n} b_{ij}^{(\sigma)} (1;-1;\sigma-1) a_j^{1-\sigma+1} ;$$

inserting this into

$$\sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \left((k;-1;\varkappa-1) a_i^{k-\varkappa+1} \psi_{1,i}^{(0)} + \sum_{\sigma=0}^{\varkappa-2} \binom{\varkappa-1}{\sigma} (k;-1;\sigma) a_i^{k-\sigma} \psi_{1,i}^{(\varkappa-\sigma-1)} \right)$$

we get (1.74) .

5.) From (1.46) we have

$$(1.77) \quad \psi_i^{(\varkappa-\sigma)} = \sum_{\substack{\lambda_1+\dots+\lambda_s=\varkappa-\sigma-1 \\ \lambda_1 \geq 0}} \frac{(\varkappa-\sigma-1)!}{\lambda_1!\dots\lambda_s!} \psi_{1,i}^{(\lambda_1)} \dots \psi_{s,i}^{(\lambda_s)} .$$

Because of (1.76) in the case of $[\phi_1\dots\phi_s\phi^k]$ the summation in (1.46)

is to extend over the following tupels

$$(\lambda_1,\dots,\lambda_s,\dots,\lambda_{s+k}) = (\lambda_1,\dots,\lambda_s,\underbrace{1,\dots,1}_{\sigma},\underbrace{0,\dots,0}_{k-\sigma})$$

with

$$\lambda_i \geqslant 0 \ (i=1,\dots,s)\ ,\ \sum_{i=1}^{s}\lambda_i + \sigma = \varkappa-1\ ,\ \sigma = 0,\dots,\varkappa-1\ .$$

Then

$$[\phi_1\dots\phi_s\phi^k] = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\substack{\lambda_1+\dots+\lambda_s-\sigma=\varkappa-1\\ \lambda_1\geqslant 0\ ,\ \sigma\geqslant 0}} \sum_{(\underbrace{1,\dots,1}_{\sigma},\underbrace{0,\dots,0}_{k-\sigma})} \frac{(\varkappa-1)!}{\lambda_1!\dots\lambda_s!}$$

$$\cdot\, \psi_{1,i}^{(\lambda_1)}\dots\psi_{s,i}^{(\lambda_s)} a_i^{k-\sigma} =$$

$$\sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1} \sum_{\substack{\lambda_1+\dots+\lambda_s=\varkappa-\sigma-1\\ \lambda_1\geqslant 0}} \binom{k}{\sigma}\frac{(\varkappa-1)!}{\lambda_1!\dots\lambda_s!}\psi_{1,i}^{(\lambda_1)}\dots\psi_{s,i}^{(\lambda_s)} a_i^{k-\sigma} =$$

$$\sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1} \binom{\varkappa-1}{\sigma}(k;-1;\sigma) a_i^{k-\sigma} \sum_{\substack{\lambda_1+\dots+\lambda_s=\varkappa-\sigma-1\\ \lambda_1\geqslant 0}} \frac{(\varkappa-\sigma-1)!}{\lambda_1!\dots\lambda_s!}\cdot$$

$$\cdot\,\psi_{1,i}^{(\lambda_1)}\dots\psi_{s,i}^{(\lambda_s)} \qquad (1.77)$$

$$\sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} \sum_{\sigma=0}^{\varkappa-1} \binom{\varkappa-1}{\sigma}(k;-1;\sigma) a_i^{k-\sigma}\psi_i^{(\varkappa-\sigma)} \quad .$$

V.2. Implicit Runge-Kutta-Processes with multiple Nodes

Introduction

As with classical Runge-Kutta methods, also here the following distinctions are useful:

explicit method: $b_{ij}^{(k)}=0$ for $j=i,i+1,\dots,n$; $k=1,2,\dots,m$

semiexplicit method: $b_{ij}^{(k)}=0$ for $j=i+1,\dots,n$; $k=1,2,\dots,m$

implicit method: otherwise .

In the first case the values $g_i^{(k)}$ can be evaluated recursively. Otherwise they are determined by implicite equations which may be solved by iterations.

The following theorem about implicite Runge-Kutta-processes is due to Butcher /2 / :

Theorem: Each quadrature formula (with single nodes)

$$(2.1)\qquad \bar{y}(x) = y_0+h\sum_{i=1}^{n} c_i f(x_0+a_i h)$$

can be extended to a implicite Runge-Kutta-process with the same order:

$$y(x)= y_0+h\sum_{i=1}^{n} c_i g_i \quad \text{with} \quad g_i= f(x_0+a_i h, y_0+h\sum_{j=1}^{n} b_{ij} g_j) \ ,$$

where a_i , c_i are the values of (2.1) and b_{ij} are determined by the equations

$$\sum_{j=1}^{n} b_{ij} a_j^{k-1} = \frac{a_i^k}{k} \qquad k=1,\dots,n \ ; \ i=1,\dots,n \ .$$

Here we are showing that an analogous theorem is also valid for quadrature formulas with multiple nodes.

Quadrature Formulas with multiple Nodes

The following generalization of (2.1)

$$(2.2)\quad \bar{y}(x)= y_0+h\sum_{i=1}^{n} c_i^{(1)} f(x_0+a_i h)+h^2\sum_{i=1}^{n} c_i^{(2)} f'(x_0+a_i h)+\ldots$$
$$\ldots+h^m\sum_{i=1}^{n} c_i^{(m)} f^{(m-1)}(x_0+a_i h)$$

is called a quadrature formulas with multiple nodes. Here not only the values of $f(x_0+a_i h)$, but also derivatives of it are evaluated.

Such formulas (or special cases) have been investigated by a number of authors, e.g. D.D.Stancu, A.H.Stroud (/48/, /50/) , S.Filippi /15/ .... Here we restrict ourselfes to the formulas with multiple Gaussian nodes given in Stroud - Stancu /50/. These reach, similiar to classical Gaussian formulas, the highest possible order.

The following theorem is proved in /50/:

Theorem: If m is odd, the coefficients a_i can be determined so that formula (2.2) reaches order $(m+1)n$, where the coefficients $c_i^{(k)}$ $(k=1,\ldots,m)$ are given by:

$$(2.3)\quad \sum_{k=1}^{m}\sum_{i=1}^{n} c_i^{(k)}(l-1;-1;k-1)a_i^{l-1} = \frac{1}{l} \quad l=1,\ldots,(m+1)n .$$

These coefficients are tabulated to 20 D in /50/ for m=3,5 and n=2(1)7 (for the interval $[-1,+1]$) .

Implicite Runge-Kutta-Process with multiple Nodes

The idea of the following proof is due to Butcher / 2/. The verification of the single steps, however, here is very muchmore complicated.

We first prove the following formula:

$$(2.4)\quad (r_1+\ldots+r_s;-1;\mu)=\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu\\ \kappa_i\geq 0}}\frac{\mu!}{\kappa_1!\ldots\kappa_s!}(r_1;-1;\kappa_1)\ldots(r_s;-1;\kappa_s)$$

where

$$(2.4a)\quad (r;-1;\mu)=r(r-1)\ldots(r-\mu+1)\ (\mu=1,2,\ldots)\ ;\ (r;-1;0)=1$$

Proof (by induction on μ) :

the case $\mu=0$ is correct.

Induction from μ to $\mu+1$:

$$(r_1+\ldots+r_s;-1;\mu+1)=(r_1+\ldots+r_s;-1;\mu).(r_1+\ldots+r_s-\mu)=$$

$$\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu\\ \kappa_1\geq 0}}\frac{\mu!}{\kappa_1!\ldots\kappa_s!}(r_1;-1;\kappa_1)\ldots(r_s;-1;\kappa_s).\left[(r_1-\kappa_1)+\ldots+(r_s-\kappa_s)\right]=$$

$$\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu\\ \kappa_1\geq 0}}\frac{\mu!}{\kappa_1!\ldots\kappa_s!}\sum_{i=1}^{s}(r_1;-1;\kappa_1)\ldots(r_s;-1;\kappa_s)(r_i-\kappa_i)=$$

$$\sum_{i=1}^{s}\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu\\ \kappa_1\geq 0}}\frac{\mu!}{\kappa_1!\ldots\kappa_s!}(r_1;-1;\kappa_1)\ldots(r_i;-1;\kappa_i+1)\ldots(r_s;-1;\kappa_s)=$$

$$\sum_{i=1}^{s}\sum_{\substack{\kappa_1+\ldots+(\kappa_i+1)+\ldots+\kappa_s=\mu+1\\ \kappa_1\geq 0}}\frac{\mu!(\kappa_i+1)}{\kappa_1!\ldots(\kappa_i+1)!\ldots\kappa_s!}(r_1;-1;\kappa_1)\ldots(r_i;-1;\kappa_i+1)\ldots$$

$$\ldots(r_s;-1;\kappa_s)\qquad(\kappa_i+1 \text{ =:> } \kappa_i)$$

$$\sum_{i=1}^{s}\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu+1\\ \kappa_1\geq 0}}\frac{\mu!\,\kappa_i}{\kappa_1!\ldots\kappa_s!}(r_1;-1;\kappa_1)\ldots(r_s;-1;\kappa_s)=$$

$$\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu+1\\ \kappa_1\geq 0}}\frac{\mu!}{\kappa_1!\ldots\kappa_s!}\left(\sum_{i=1}^{s}\kappa_i\right)(r_1;-1;\kappa_1)\ldots(r_s;-1;\kappa_s)=$$

$$\sum_{\substack{\kappa_1+\ldots+\kappa_s=\mu+1\\ \kappa_1\geq 0}}\frac{(\mu+1)!}{\kappa_1!\ldots\kappa_s!}(r_1;-1;\kappa_1)\ldots(r_s;-1;\kappa_s)\ .\qquad\text{Done.}$$

Consequence:

$$(2.5a)\quad (k+l-1;-1;s)=(k-1;-1;s)+l\sum_{\kappa=0}^{s-1}\binom{s}{\kappa+1}(l-1;-1;\kappa)(k-1;-1;s-\kappa-1)$$

$$(2.5b)\qquad =(l-1;-1;s)+k\sum_{\kappa=0}^{s-1}\binom{s}{\kappa+1}(k-1;-1;\kappa)(l-1;-1;s-\kappa-1)$$

(2.5a) follows from (2.4) for $r_1=k-1$, $r_2=1$, $\mu=s$; (2.5b) is (2.5a) with k and l interchanged.

For the rest of this section we assume the following conditions:

$$(2.6)\qquad \begin{aligned} &a_i \neq a_k \quad \text{for } i \neq k \quad \text{and } a_i \neq 0 \quad (i=1,\ldots,n)\\ &c_i^{(k)} \neq 0 \quad i=1,\ldots,n \quad , \quad k=1,\ldots,m \quad . \end{aligned}$$

Using (1.72) , (1.75) we now define the following symbols:

Defintion:

$$A(\xi) \Longleftrightarrow \gamma\phi = 1 \quad \text{for all elementary weight of order } r$$

$$B(\xi) \Longleftrightarrow [\phi^{k-1}] = \sum_{\kappa=1}^{m}\sum_{i=1}^{n} c_i^{(\kappa)}(k-1;-1;\kappa-1)a_i^{k-\kappa} = \frac{1}{k} \qquad k=1,\ldots,\xi$$

$$C(\xi) \Longleftrightarrow \sum_{\kappa=1}^{m}\frac{(k-1;-1;\kappa-1)}{\kappa!}\sum_{j=1}^{n} b_{ij}^{(\kappa)}a_j^{k-\kappa} = \frac{a_i^k}{k} \qquad \begin{matrix} i=1,\ldots,n\\ k=1,\ldots,\xi \end{matrix}$$

$$D(\xi) \Longleftrightarrow \sum_{\kappa=1}^{m}\sum_{i=1}^{n} c_i^{(\kappa)}(k-1;-1;\kappa-1)a_i^{k-\kappa}b_{ij}^{(\sigma)} =$$

$$\sigma!\left(\frac{c_j^{(\sigma)}(1-a_i^k)}{k} - \sum_{\kappa=\sigma+1}^{m} c_j^{(\kappa)}\binom{\kappa}{\sigma}(k-1;-1;\kappa-\sigma-1)a_j^{k+\sigma-\kappa}\right)$$

$$\sigma=1,\ldots,m \quad j=1,\ldots,n \quad k=1,\ldots,\xi$$

$$E(\xi,\eta) \Longleftrightarrow [\phi^{k-1}[\phi^{l-1}]] = \sum_{\kappa=1}^{m}\sum_{i=1}^{n} c_i^{(\kappa)}a_i^{k-\kappa}\left(\sum_{\sigma=1}^{\kappa}\frac{(l-1;-1;\sigma-1)(k-1;-1;\kappa-1)}{\sigma!}\right.$$

$$\left.\cdot\sum_{j=1}^{n} b_{ij}^{(\sigma)}a_j^{l-\sigma}+a_i^{l}\sum_{\sigma=0}^{\kappa-2}\binom{\kappa-1}{\sigma+1}(k-1;-1;\kappa-\sigma-2)(l-1;-1;\sigma)\right) = \frac{1}{l(k+l)}$$

$$l=1,\ldots,n \qquad k=1,\ldots,\xi$$

Theorem:

$$(2.7)\qquad B(\xi+\eta)\ ,\ C(\eta) \quad \Longrightarrow \quad E(\xi,\eta) \quad .$$

Proof:

$$\sum_{\varkappa=1}^{n}\sum_{i=1}^{n} c_i^{(\varkappa)} a_i^{k-\varkappa}\left(\sum_{\sigma=1}^{l}\frac{(l-1;-1;\sigma-1)(k-1;-1;\varkappa-1)}{\sigma!}\sum_{j=1}^{n} b_{ij}^{(\sigma)} a_j^{l-\sigma} + \right.$$

$$\left. + a_i^{l}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(l-1;-1;\sigma)\right) \overset{(C(\eta))}{=}$$

$$\sum_{\varkappa=1}^{n}\sum_{i=1}^{n} c_i^{(\varkappa)} a_i^{k-\varkappa+l}\cdot\frac{1}{l}\left((k-1;-1;\varkappa-1)+l\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(l-1;-1;\sigma)\right) \quad (2.5)$$

$$\frac{1}{l}\sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)}(k+l-1;-1;\varkappa-1)a_i^{k+l-\varkappa}\, B(\xi+\eta)\ \frac{1}{l}\cdot\frac{1}{l+k}\quad .$$

Theorem:

$$(2.8)\qquad B(\xi + m.n)\ ,\ E(\xi\ ,\ m.n) \Longrightarrow D(\xi)\quad .$$

Proof:

$$\frac{1}{l(l+k)} = \frac{1}{k}\cdot\left(\frac{1}{l}-\frac{1}{l+k}\right)B(\xi+\eta)$$

$$= \frac{1}{k}\sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} a_j^{l-\varkappa}\left((l-1;-1;\varkappa-1)-(k+l-1;-1;\varkappa-1)a_j^{k}\right) \qquad (2.5b)$$

$$\sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} a_j^{l-\varkappa}\left((l-1;-1;\varkappa-1)\frac{1-a_j^{k}}{k} - a_j^{k}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\sigma)(l-1;-1;\varkappa-\sigma-2)\right)$$

$$(\sigma \longrightarrow \varkappa-\sigma-2)$$

$$\sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} a_j^{l-\varkappa}\left((l-1;-1;\varkappa-1)\frac{1-a_j^{k}}{k} - a_j^{k}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma}(k-1;-1;\varkappa-\sigma-2)(l-1;-1;\sigma)\right) =$$

with $\binom{\varkappa-1}{\sigma} = \binom{\varkappa}{\sigma+1} - \binom{\varkappa-1}{\sigma+1}$

$$= \sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} a_j^{l-\varkappa}\left((l-1;-1;\varkappa-1)\frac{1-a_j^{k}}{k} + a_j^{k}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(l-1;-1;\sigma)\right) -$$

$$- \sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_i^{(\varkappa)} a_j^{k+l-\varkappa}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(l-1;-1;\sigma) =$$

(in the last expression we now substitute $\sigma \dashrightarrow \varkappa-1$ and $\varkappa \dashrightarrow \sigma$: the summation bounds $0 \leq \sigma \leq \varkappa-2 \leq m-2$ thus transform into $0 \leq \varkappa-1 \leq \sigma-2 \leq m-2$, hence $1 \leq \varkappa \leq m-1$, $\varkappa+1 \leq \sigma \leq m$.)

$$= \sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} a_j^{1-\varkappa}\left((1-1;-1;\varkappa-1)\,\frac{1-a_j^k}{k} + a_j^k \sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(1-1;-1;\sigma)\right)$$

$$- \sum_{\varkappa=1}^{m}\sum_{\sigma=\varkappa+1}^{m+1}\sum_{j=1}^{n} c_j^{(\sigma)} a_j^{k+1-\sigma}\binom{\sigma}{\varkappa}(k-1;-1;\sigma-\varkappa-1)(1-1;-1;\varkappa-1) =$$

(by rearranging)

$$\sum_{\varkappa=1}^{m}\sum_{j=1}^{n}(1-1;-1;\varkappa-1)a_j^{1-\varkappa}\left(c_j^{(\varkappa)}\,\frac{1-a_j^k}{k} - \sum_{\sigma=\varkappa+1}^{m} c_j^{(\sigma)}\binom{\sigma}{\varkappa}(k-1;-1;\sigma-\varkappa-1)a_j^{k+\varkappa-\sigma}\right) +$$

$$(2.9)\qquad + \sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} a_j^{k+1-\varkappa}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(1-1;-1;\sigma) \quad ;$$

on the other side we have from $E(\xi,\eta)$ after slight modifications:

$$\frac{1}{1(1+k)}\,E(\xi,\eta)\sum_{\varkappa=1}^{m}\sum_{j=1}^{n}(1-1;-1;\varkappa-1)a_j^{1-\varkappa}\cdot\frac{1}{\varkappa!}\left(\sum_{\sigma=1}^{m}\sum_{i=1}^{n} c_i^{(\sigma)}(k-1;-1;\sigma-1)a_i^{k-\sigma}b_{ij}^{(\varkappa)}\right)$$

$$(2.10)\qquad + \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)} a_i^{k+1-\varkappa}\sum_{\sigma=0}^{\varkappa-2}\binom{\varkappa-1}{\sigma+1}(k-1;-1;\varkappa-\sigma-2)(1-1;-1;\sigma)\ .$$

Subtracting (2.10) and (2.9) we obtain

$$\sum_{\varkappa=1}^{m}\sum_{j=1}^{n}(1-1;-1;\varkappa-1)a_j^{1-\varkappa}\cdot\left\{\frac{1}{\varkappa!}\sum_{\sigma=1}^{m}\sum_{i=1}^{n} c_i^{(\sigma)}(k-1;-1;\sigma-1)a_i^{k-\sigma}b_{ij}^{(\varkappa)} - \right.$$

$$\left. - \left(c_j^{(\varkappa)}\,\frac{1-a_j^k}{k} - \sum_{\sigma=\varkappa+1}^{m} c_j^{(\sigma)}\binom{\sigma}{\varkappa}(k-1;-1;\sigma-\varkappa-1)a_j^{k+\varkappa-\sigma}\right)\right\} = 0\ .$$

Now we consider the expressions in the waved brackets $\{\ldots\}$ as independent variables and let l run from 1 to m.n , then this is a homogeneous linear system of equations with non- zero determinant (cf. footnote on p.103 and (2.6)). Hence the system has only the zero-solution $\{\ldots\} = 0$ $(k=1,\ldots,m\ ;\ j=1,\ldots,n\ ;\ k=1,\ldots,\xi)$, this means that $D(\xi)$ is satisfied. Done.

Defintion: Given $\phi_1 = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)}\psi_{1,i}^{(\varkappa)}$ and $\phi_2 = \sum_{\varkappa=1}^{m}\sum_{i=1}^{n} c_i^{(\varkappa)}\psi_{2,i}^{(\varkappa)}$;

then we write

$$(2.11)\qquad \phi_1 \equiv \phi_2 \iff \psi_{1,i}^{(\varkappa)} = \psi_{2,i}^{(\varkappa)}\ (i=1,\ldots,n\ ;\ \varkappa=1,\ldots,m)\ .$$

(1.46) shows that $\phi_i \equiv \hat{\phi}_i$ $(i=1,\dots,s)$ yields $[\phi_1\dots\phi_s] \equiv [\hat{\phi}_1\dots\hat{\phi}_s]$.

Lemma 1: If $C(\eta)$ holds and $\phi = [\phi_1\dots\phi_s]$ is such that ϕ_i have orders $r_i \leq \eta$ $(i=1,\dots,s)$, then

$$\gamma\phi = r\,[\phi^{r-1}] \,. \tag{2.12}$$

Proof: (by induction on r):

for r=2 (2.12) is satisfied , since $\phi = [\phi]$ is the only elementary weight of that order.

First $C(\eta)$ gives since $r_i \leq \eta$ $(i=1,\dots,s)$:

$$\hat{\phi} \equiv [[\phi^{r_1-1}]\dots[\phi^{r_s-1}]] \equiv \frac{1}{r_1\dots r_s}\,[\phi^{r-1}] \text{ with } r=r_1+\dots+r_s+1 \,; \tag{2.13}$$

because:

the coefficients $\hat{\Psi}^{(k)}_{i,j}$ belonging to $[\phi^{r_i-1}]$ read as follows:

$$\hat{\Psi}^{(k)}_{i,j} = (r_i-1;-1;k-1)a_j^{r_i-k} = \frac{(r_i;-1;k)}{r_i}\,a_j^{r_i-k} \quad (k=0,1,\dots,m) \,. \tag{2.14}$$

For k=1,...,m this follows immediately from (1.72) and for k=0 from

$$\hat{\Psi}^{(0)}_{i,j} = \sum_{\kappa=1}^{m}\frac{1}{\kappa!}\sum_{k=1}^{m} b^{(\kappa)}_{jk}\hat{\Psi}^{(\kappa)}_{i,k} = \sum_{\kappa=1}^{m}\frac{1}{\kappa!}\sum_{k=1}^{m} b^{(\kappa)}_{jk}(r_i;-1;\kappa-1)a_k^{r_i-\kappa}\quad C(\eta)\,\frac{a_j^{r_i}}{r_i} \,.$$

Thus

$$\hat{\phi} \overset{(1.46)}{=} \sum_{\kappa=1}^{m}\sum_{j=1}^{n} c^{(\kappa)}_j \sum_{\substack{\sigma_1+\dots+\sigma_s=\kappa-1\\ \sigma_i\geq 0}} \frac{(\kappa-1)!}{\sigma_1!\dots\sigma_s!}\,\hat{\Psi}^{(\sigma_1)}_{1,j}\dots\hat{\Psi}^{(\sigma_s)}_{s,j} \overset{(2.14)}{=}$$

$$\sum_{\kappa=1}^{m}\sum_{j=1}^{n} c^{(\kappa)}_j \sum_{\substack{\sigma_1+\dots+\sigma_s=\kappa-1\\ \sigma_i\geq 0}} \frac{(\kappa-1)!}{\sigma_1!\dots\sigma_s!}\cdot\frac{(r_1;-1;\sigma_1)}{r_1}\dots\frac{(r_s;-1;\sigma_s)}{r_s}\cdot a_j^{r_1+\dots+r_s-(\kappa-1)} \overset{(2.13)}{=}$$

$$\sum_{\kappa=1}^{m}\sum_{j=1}^{n} c^{(\kappa)}_j \frac{a_j^{r-\kappa}}{r_1\dots r_s} \sum_{\substack{\sigma_1+\dots+\sigma_s=\kappa-1\\ \sigma_i\geq 0}} \frac{(\kappa-1)!}{\sigma_1!\dots\sigma_s!}\,(r_1;-1;\sigma_1)\dots(r_s;-1;\sigma_s)$$

(using (2.4))

$$\frac{1}{r_1\dots r_s}\sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)}(r-1;-1;\varkappa-1)a_j^{r-\varkappa} \quad (1.72) \quad \frac{1}{r_1\dots r_s}[\phi^{r-1}] \ .$$

Induction from r-1 to r :

Induction hypothesis:

$$(2.15) \qquad \gamma_i\phi_i \equiv r_i[\phi^{r_i-1}] \text{ where } r_i \leq \eta \ \ (i=1,\dots,s) \ ;$$

thus we have

$$\gamma\phi \equiv r\gamma_1\dots\gamma_s[\phi_1\dots\phi_s] \overset{(2.15)}{\equiv} r.r_1\dots r_s[[\phi^{r_1-1}]\dots[\phi^{r_s-1}]] \overset{(2.13)}{\equiv} r[\phi^{r-1}].$$

Done.

Corollary: If $C(\eta)$ is satisfied, then (2.12) is valid for all elementary weights of order $r \leq \eta+1$.

Since for these the conditions of Lemma are satisfied.

Lemma 2: If $C(\eta)$ is valid and $\phi = [\phi_1\dots\phi_s]$ is such that for the corresponding orders it holds that $r_1 \geq \eta+1$, $r_i \leq \eta$ $(i=2,\dots,s)$, then

$$(2.16) \qquad \gamma\phi \equiv r\gamma_1[\phi_1\phi^{r-r_1-1}] \ .$$

Proof:

First we have by Lemma 1 and $C(\eta)$

$$(2.17) \qquad [\phi_1[\phi^{r_2-1}]\dots[\phi^{r_s-1}]] \equiv \frac{1}{r_2\dots r_s}[\phi_1\phi^{r_2+\dots+r_s}]$$

where $r_1 > \eta$, $r_i \leq \eta$ $(i=2,\dots,s)$;

since

$$[\phi_1[\phi^{r_2-1}]\dots[\phi^{r_s-1}]] \overset{(1.46)}{\equiv} \sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} \sum_{\substack{\sigma_1+\dots+\sigma_s=\varkappa-1\\ \sigma_i\geq 0}} \frac{(\varkappa-1)!}{\sigma_1!\dots\sigma_s!}\Psi_{1,j}^{(\sigma_1)}\dots\Psi_{s,j}^{(\sigma_s)}$$

(using (2.14))

$$\sum_{\varkappa=1}^{m}\sum_{j=1}^{n} c_j^{(\varkappa)} \sum_{\substack{\sigma_1+\dots+\sigma_s=\varkappa-1\\ \sigma_i\geq 0}} \frac{(\varkappa-1)!}{\sigma_1!\dots\sigma_s!}\Psi_{1,j}^{(\sigma_1)} \frac{(r_2;-1;\sigma_2)\dots(r_s;-1;\sigma_s)}{r_2\dots r_s} \cdot$$

$$\cdot a_j^{r_2+\dots+r_s-(\varkappa-\sigma_1-1)} \equiv$$

$$\frac{1}{r_2\cdots r_s}\sum_{\varkappa=1}^{m}\sum_{j=1}^{n}c_j^{(\varkappa)}\sum_{\sigma_1=0}^{\varkappa-1}\frac{(\varkappa-1)!}{\sigma_1!(\varkappa-\sigma_1-1)!}\sum_{\substack{\sigma_2+\ldots+\sigma_s=\varkappa-\sigma_1-1\\ \sigma_i\geq 0}}\frac{(\varkappa-\sigma_1-1)!}{\sigma_2!\ldots\sigma_s!}\cdot$$

$$\cdot(r_2;-1;\sigma_2)\ldots(r_s;-1;\sigma_s)a_j^{r_2+\ldots+r_s-(\varkappa-\sigma_1-1)} \equiv$$

(using (2.4))

$$\frac{1}{r_2\cdots r_s}\sum_{\varkappa=1}^{m}\sum_{j=1}^{n}c_j^{(\varkappa)}\sum_{\sigma_1=0}^{\varkappa-1}\binom{\varkappa-1}{\sigma_1}(r_2+\ldots+r_s;-1;\varkappa-\sigma_1-1)a_j^{r_2+\ldots+r_s-(\varkappa-\sigma_1-1)}\Psi_{1,j}^{(\sigma_1}$$

($\sigma_1 \dashrightarrow \varkappa-\sigma_1-1$)

$$\frac{1}{r_2\cdots r_s}\sum_{\varkappa=1}^{m}\sum_{j=1}^{n}c_j^{(\varkappa)}\sum_{\sigma=0}^{\varkappa-1}\binom{\varkappa-1}{\sigma}(r_2+\ldots+r_s;-1;\sigma)a_j^{r_2+\ldots+r_s-\sigma}\Psi_{1,j}^{(\varkappa-\sigma-1)} \quad (1.73)$$

$$\equiv \frac{1}{r_2\cdots r_s}\left[\phi_1\phi^{r_2+\ldots+r_s}\right] .$$

Thus

$$\gamma\phi \equiv r\gamma_1\cdots\gamma_s\left[\phi_1\ldots\phi_s\right] \overset{(2.12)}{\equiv} r\gamma_1 r_2\cdots r_s\left[\phi_1\left[\phi^{r_2-1}\right]\ldots\left[\phi^{r_s-1}\right]\right] \overset{(2.17)}{\equiv}$$

$$r\gamma_1\left[\phi_1\phi^{r_2+\ldots+r_s}\right] .$$

Done.

Theorem:

$$\left.\begin{array}{l} B(\xi)\ ,\ C(\eta)\ ,\ D(\xi) \\ \xi \leq \xi+\eta+1\ ,\ \xi \leq 2\eta+2 \end{array}\right\} \Rightarrow A(\xi)\ ; \qquad (2.18)$$

i.e., the Runge-Kutta-process has the order ξ .

Proof:

1.) If $\phi = [\phi_1\ldots\phi_s]$ with $r_i \leq \eta$ $(i=1,\ldots,s)$ and $r \leq \xi$ we have by Lemma 1 and $B(\xi)$

$$\gamma\phi \overset{(2.12)}{\equiv} r\left[\phi^{r-1}\right] B(\xi) \quad 1 \quad .$$

In particular this is the case if $r \leq \eta+1$ (Corollary to Lemma 1).

2.) Thus it sufficients to consider the elementary weights $\phi = [\phi_1\ldots\phi_s]$ for which at least <u>one</u> ϕ_i , say ϕ_1 , has order

$r_1 > \eta$.

From the condition $\xi \leq 2\eta+2$ it now follows that the others elementary weights ϕ_i have orders $r_i \leq \eta$ $(i=2,\ldots,s)$, otherwise, if for example $r_2 > \eta$, then we have

$$\xi \geq r = r_1+r_2+\ldots+r_s+1 \geq (\eta+1)+(\eta+1)+1 = 2\eta+3 > \xi \quad ,$$

hence a contradiction.

Thus the remaining elementary weights ϕ satisfies the conditions of Lemma 2, and it is suffient to show that

$$(2.19) \quad [\phi_1 \phi^{r-r_1-1}] = \frac{1}{r\gamma_1} \quad ;$$

since then we have

$$\gamma\phi \overset{(2.16)}{=} r\gamma_1 [\phi_1 \phi^{r-r_1-1}] \overset{(2.19)}{=} 1 \ .$$

Proof of (2.19):

This proof is by double induction over r and r_1, for $r = \eta+2,\ldots,\xi$ and $r_1 = \eta+1,\ldots,r-1$.

First we have from the conditions (2.18) and because

$$(2.20) \quad r-r_1 \leq \xi+\eta+1-(\eta+1) = \xi \quad ;$$

and hence

$$[\phi_1 \phi^{r-r_1-1}] \overset{(1.73)}{=} \sum_{\kappa=1}^{m} \sum_{i=1}^{n} c_i^{(\kappa)} \Big((r-r_1-1;-1;\kappa-1) a_i^{r-r_1-\kappa} \psi_{1,i}^{(o)} +$$

$$+ \sum_{\sigma=o}^{\kappa-2} \binom{\kappa-1}{\sigma} (r-r_1-1;-1;\sigma) a_i^{r-r_1-\sigma-1} \psi_{1,i}^{(\kappa-\sigma-1)} \Big) \overset{(1.36)}{=}$$

$$\sum_{\kappa=1}^{m} \sum_{i=1}^{n} c_i^{(\kappa)} \Big((r-r_1-1;-1;\kappa-1) a_i^{r-r_1-\kappa} \sum_{\sigma=1}^{m} \frac{1}{\sigma!} \sum_{k=1}^{n} b_{ik}^{(\sigma)} \psi_{1,k}^{(\sigma)} +$$

$$+ \sum_{\sigma=o}^{\kappa-2} \binom{\kappa-1}{\sigma} (r-r_1-1;-1;\sigma) a_i^{r-r_1-\sigma-1} \psi_{1,i}^{(\kappa-\sigma-1)} \Big) =$$

(interchange of sequence of summation in the first expression, substitution $\sigma \longrightarrow \kappa-\sigma-1$, $i \longrightarrow k$ in the second)

$$\sum_{\sigma=1}^{m} \sum_{k=1}^{n} \Big(\frac{1}{\sigma!} \sum_{\kappa=1}^{m} \sum_{i=1}^{n} c_i^{(\kappa)} (r-r_1-1;-1;\kappa-1) a_i^{r-r_1-\kappa} b_{ik}^{(\sigma)} \Big) \psi_{1,k}^{(\sigma)} +$$

$$\sum_{\kappa=1}^{m} \sum_{k=1}^{n} c_k^{(\kappa)} \sum_{\sigma=1}^{\kappa-1} \binom{\kappa-1}{\sigma} (r-r_1-1;-1;\kappa-\sigma-1) a_k^{r-r_1-\kappa+\sigma} \psi_{1,k}^{(\kappa)} \overset{(2.20) \text{ and } D(\xi)}{=}$$

$$\sum_{\sigma=1}^{m}\sum_{k=1}^{n}\left(\frac{c_k^{(\sigma)}(1-a_k^{r-r_1})}{r-r_1} - \sum_{\kappa=\sigma+1}^{m} c_k^{(\kappa)}\binom{\kappa}{\sigma}(r-r_1-1;-1;\kappa-\sigma-1)a_k^{r-r_1-\kappa+\sigma}\right)\Psi_{1,k}^{(\sigma)}$$

$$+\sum_{\kappa=1}^{m}\sum_{k=1}^{n} c_k^{(\kappa)}\sum_{\sigma=1}^{\kappa-1}\binom{\kappa-1}{\sigma}(r-r_1-1;-1;\kappa-\sigma-1)a_k^{r-r_1-\kappa+\sigma}\Psi_{1,k}^{(\kappa)} = \left(\text{with } \binom{\kappa}{\sigma}=\binom{\kappa-1}{\sigma}+\binom{\kappa-1}{\sigma-1}\right.$$

$$\frac{1}{r-r_1}\sum_{\sigma=1}^{m}\sum_{k=1}^{n} c_k^{(\sigma)}\Psi_{1,k}^{(\sigma)} -$$

$$\frac{1}{r-r_1}\sum_{\kappa=1}^{m}\sum_{k=1}^{n} c_k^{(\kappa)}\left(a_k^{r-r_1}\Psi_{1,k}^{(\kappa)}+(r-r_1)\sum_{\sigma=1}^{\kappa-1}\binom{\kappa-1}{\sigma-1}(r-r_1-1;-1;\kappa-\sigma-1)a_k^{r-r_1-\kappa+\sigma}\Psi_{1,k}^{(\sigma)}\right)$$

$$(\ \sigma \longrightarrow \kappa-\sigma\ )$$

$$\frac{1}{r-r_1}\phi_1 - \frac{1}{r-r_1}\sum_{\kappa=1}^{m}\sum_{k=1}^{n} c_k^{(\kappa)}\sum_{\sigma=0}^{\kappa-1}\binom{\kappa-1}{\sigma}(r-r_1;-1;\sigma)a_k^{r-r_1-\sigma}\Psi_{1,k}^{(\kappa-\sigma)} \qquad (1.75)$$

$$\frac{1}{r-r_1}\left(\phi_1 - [\hat{\phi}_1\ldots\hat{\phi}_t\phi^{r-r_1}]\right)$$

with $\phi_1 = [\hat{\phi}_1\ldots\hat{\phi}_t]$.

Since $r_1 < r$ we have the induction hypothesis[1] $\phi_1 = \frac{1}{\gamma_1}$ and since the highest order of $\phi_1,\ldots,\phi_t$ is smaller than r_1, we may also assume the induction hypothesis for r_1 ; this yields

$$[\hat{\phi}_1\ldots\hat{\phi}_t\phi^{r-r_1}] = \frac{1}{r\gamma_1\ldots\gamma_t} = \frac{r_1}{r\gamma_1} \quad .$$

This gives

$$[\phi_1\phi^{r-r_1-1}] = \frac{1}{r-r_1}\left(\frac{1}{\gamma_1} - \frac{1}{r\gamma_1}\right) = \frac{1}{r\gamma_1} \quad , \text{ hence (2.19). Done.}$$

Theorem: it holds that

(2.22) $\quad B(m.n+n)\ ,\ C(m.n) \Longrightarrow A(m.n+n)$.

Proof:

$\xi = n\ ,\ \eta = m.n\ ,\ \zeta = m.n+n$

$B(m.n+n)\ ,\ C(m.n) \xrightarrow{(2.7)} E(n,m.n)$;

---

[1] The induction start with $r = \eta+2$, $r_1 = \eta+1$ where (2.19) is correct since then the elementary weights in (2.21a) satisfy Lemma 1.

$B(n+m.n)$, $E(n,m.n) \xrightarrow{(2.8)} D(n)$;

since $\xi \leq \xi + \eta + 1$ and $\xi \leq 2\eta + 2$ are valid we have

$B(n+m.n)$, $C(m.n)$, $D(n) \xrightarrow{(2.18)} A(n+m.n)$. Done.

We are now ready to formulate our main theorem:

Theorem: The quadrature formula with multiple Gaussian nodes (2.2) can be extended to an implicit Runge-Kutta-process with multiple nodes. This has the same order than the quadrature formula.
The coefficients $b_{ij}^{(k)}$ are uniquely determined by the nodes a_i and the weights $c_i^{(k)}$.

Proof:

$B(m.n+n)$ is satisfied because of (2.3).

By

$$C(m.n) \Longleftrightarrow \sum_{\varkappa=1}^{m} \frac{(k-1;-1;\varkappa-1)}{\varkappa!} \sum_{j=1}^{n} b_{ij}^{(\varkappa)} a_j^{k-\varkappa} = \frac{a_i^k}{k} \qquad \begin{matrix} k=1,\dots,m.n \\ i=1,\dots,n \end{matrix}$$

the coefficients $b_{ij}^{(k)}$ are uniquely determined since the corresponding determinant[1] does not vanish.
Finally from (2.22) we have that the Runge-Kutta-process has order $(m.n+n)$. Done.

We still remark that all other theorems of Butcher can be generalized. We state them without proof since we do not need them:

Theorem: $B(n+\eta)$, $E(n,\eta) \Rightarrow C(\eta)$;

Theorem: $B(\xi+\eta)$, $D(\xi) \Rightarrow E(\xi,\eta)$.

---

[1] this is a so-called "confluent" Vandermonde determinant which is regular if and only if the nodes a_i are all different (cf. Gautschi /17/). But this is assured by (2.6).

The iterative Computation of the $g_i^{(k)}$

We now show that the values $g_i^{(k)}$ which are determined by the implicit system of functions

$$(2.23)\quad g_i^{(k)} = (D^k y)(x_o+a_i h, y_o+h\sum_{j=1}^{n} b_{ij}^{(1)} g_j^{(1)}+\ldots+\frac{h^m}{m!}\sum_{j=1}^{n} b_{ij}^{(m)} g_j^{(m)})$$

$$(k=1,\ldots,m \;;\; i=1,\ldots,n)$$

can be computed iteratively:

We put

$$B_k = \max_i \left\{\sum_{j=1}^{n} |b_{ij}^{(k)}|\right\} \quad (k=1,\ldots,m) \quad \text{and}$$

$$\|v\| = \max\{v_1,\ldots,v_n\} \text{ with } v=(v_1,\ldots,v_n) \quad (\text{vector norm}).$$

Theorem: If the functions $D^k y$ $(k=1,\ldots,m)$ satisfy a Lipschitz-condition

$$(2.24)\quad \|(D^k y)(z') - (D^k y)(z'')\| \leq L_k \|z'-z''\| \quad (k=1,\ldots,m)$$

in some domain B, and if the step size h satisfies the following conditions

$$(2.25a)\quad |h| L B_1 < 1 \quad \text{where} \quad L=L_1+\ldots+L_m$$

$$(2.25b)\quad \frac{|h|^{k-1}}{k!} B_k \leq B_1 \quad (k=2,\ldots,m)$$

then (2.23) posseses a unique solution.

Proof: Assume that there exist two solutions $g_i^{(k)}$ and $\bar{g}_i^{(k)}$: then

$$\|g_i^{(k)}-\bar{g}_i^{(k)}\| \overset{(2.24)}{\leq} L_k \|h\sum_{j=1}^{n} b_{ij}^{(1)}(g_j^{(1)}-\bar{g}_j^{(1)})+\ldots+\frac{h^m}{m!}\sum_{j=1}^{n} b_{ij}^{(m)}(g_j^{(m)}-\bar{g}_j^{(m)})\| \leq$$

$$L_k\left(|h| B_1 \max_j \|g_j^{(1)}-\bar{g}_j^{(1)}\|+\ldots+\frac{|h|^m}{m!} B_m \max_j \|g_j^{(m)}-\bar{g}_j^{(m)}\|\right) \overset{(2.25b)}{\leq}$$

$$L_k |h| B_1 \sum_{l=1}^{m} \max_j \|g_j^{(l)}-\bar{g}_j^{(l)}\| .$$

Since this is valid for all i , we have

$$\max_i \|g_i^{(k)}-\bar{g}_i^{(k)}\| \leq L_k |h| B_1 \sum_{l=1}^{m} \max_j \|g_j^{(l)}-\bar{g}_j^{(l)}\| \quad \text{and}$$

$$\sum_{k=1}^{m} \max_i \|g_i^{(k)} - \tilde{g}_i^{(k)}\| \leq L|h|B_1 \sum_{l=1}^{m} \max_j \|g_j^{(1)} - \tilde{g}_j^{(1)}\| \quad \text{with } L = L_1 + \ldots + L_m \ .$$

This, however, gives a contradiction with (2.25a) .

Theorem: Under the conditions of the preceding Theorem the following iteration converges to the solution of (2.23):

$$(2.26a) \quad g_{i,N}^{(k)} = (D^k y)(x_o + a_i h, y_o + h \sum_{j=1}^{n} b_{ij}^{(1)} g_{j,N-1}^{(1)} + \ldots + \frac{h^m}{m!} \sum_{j=1}^{n} b_{ij}^{(m)} g_{j,N-1}^{(m)}) \qquad (k=1,\ldots,m)$$

with

$$(2.26b) \quad g_{i,o}^{(k)} = o \ , \quad g_{i,1}^{(k)} = \left[D^k y\right]_o \ .$$

Proof:

With

$$(2.27) \quad K = L|h|B_1 < 1$$

we obtain analoguously

$$\|g_{i,N}^{(k)} - g_{i,N-1}^{(k)}\| \leq L_k |h| B_1 \sum_{l=1}^{m} \max_j \|g_{j,N-1}^{(1)} - g_{j,N-2}^{(1)}\| \text{ and}$$

$$(2.28) \quad \sum_{k=1}^{m} \max_i \|g_{i,N}^{(k)} - g_{i,N-1}^{(k)}\| \leq K \sum_{k=1}^{m} \max_i \|g_{i,N-1}^{(k)} - g_{i,N-2}^{(k)}\| \quad (N=2,3,\ldots)$$

Thus we have

$$\|g_{i,N}^{(k)} - g_{i,N-1}^{(k)}\| \leq L_k |h| B_1 \sum_{l=1}^{m} \max_j \|g_{j,N-1}^{(1)} - g_{j,N-2}^{(1)}\| \overset{(2.27)}{\leq}$$

$$\frac{L_k}{L} K \sum_{l=1}^{m} \max_j \|g_{j,N-1}^{(1)} - g_{j,N-2}^{(1)}\| \overset{(2.28)}{\leq}$$

$$\frac{L_k}{L} K^2 \sum_{l=1}^{m} \max_j \|g_{j,N-2}^{(1)} - g_{j,N-3}^{(1)}\| \ldots \overset{(2.26b)}{\leq}$$

$$\frac{L_k}{L} K^{N-1} \sum_{l=1}^{m} \left[D^l y\right]_o = \text{const } K^{N-1} \ .$$

$$\|g_{i,N}^{(k)} - g_i^{(k)}\| \leq \|g_{i,N}^{(k)} - g_{i,N+1}^{(k)}\| + \|g_{i,N+1}^{(k)} - g_{i,N+2}^{(k)}\| + \ldots + =$$

$$\sum_{\sigma=0}^{\infty} \| g_{i,N+\sigma}^{(k)} - g_{i,N+\sigma+1}^{(k)} \| \leq \text{const} \sum_{\sigma=0}^{\infty} K^{N+\sigma+1} = \text{const} \frac{K^{N+1}}{1-K} \quad . \quad \text{Done.}$$

Table of coefficients for m=3 , n=2,3,4.

The nodes a_i and weights $c_i^{(k)}$ are those of /48/ transformed to the intervall [o,1].

The coefficients are tabulated in the bystanding sequence.

Condition (2.25b) gives for the step size h the following restrictions :

m=3 n=2 : h < 11,9
n=3 : h < 15,3
n=4 : h < 2o,8 .

This, however, is no limitation to the practical use of these formulas.

a_1 .
a_n
$c_1^{(1)} \ldots c_n^{(1)}$
.....
$c_1^{(3)} \ldots c_n^{(3)}$
$b_{i1}^{(1)} \ldots b_{in}^{(1)}$ (i=1,...,n)
.....
$b_{i1}^{(3)} \ldots b_{in}^{(3)}$ (i=1,...,n)

Order 8

,18539443582504 5/+oo
,9146o5564174954/+oo

,5oooooooooooooo/+oo ,5oooooooooooooo/+oo
,24o72942o844974/-o1 -,24o72942o844974/-o1
,366264960671727/-o2 ,366264960671727/-o2

,2o185411583100 5/+oo -,164596800059598/-o1
,5164596800059 59/+oo ,2981458941689 94/+oo

-,223466569o8o541/-o1 ,86387873o02417/-o2
,56834671899919o/-o1 -,7o492541o77o49o/-o1

,1167396684oo997/-o1 -,215351251o65784/-o2
,2412941o15o9615/-o1 ,1o3o193o8oo2o39/-o1

Order 12

,9278o4o721111183/-o1
,5ooooooooooooo/+oo
,9o7219592788881/+oo

,2666582o296o883/+oo	,4666835940782222/+oo	,2666582o296o888/+oo
,7791166649283683/-o2	,ooooooooooooooo/+oo	-,7791166649283983/-o2
,5134356o115744o/-o3	,2765985622271983/-o2	,513435o9115744o/-o3

,1o377386543513o/+oo	-,1486824o47o3o24/-o1	,397476224629o81/-o2
,2755393520666o95/+oo	,233341797o39111/+oo	-,888114910720614/-o2
,2627834407145983/+oo	,4815518345485241/+oo	,162884317525759/+oo

-,4465910966705793/-o2	,2751964312342153/-o2	-,1o96496o6514297/-o2
,1796525975454253/-o1	-,3970498487494343/-o1	,2389292645597474/-o2
,1449583723342473/-o1	,2751964312342153/-o2	-,2oo4824426527353/-o1

,16592513959o3713/-o2	-,1992374o4611o183/-o2	,1351o5286925153/-o3
,3357509459o8696o6/-o2	,8297656866315943/-o2	-,2763979o3924422/-o3
,2945505260o19433/-o2	,1853768777974zo/-o1	,1421359151o4o92/-o2

Order 16

,5516159319722o5/-o1
,32o37o736438o558/+oo
,6794292635194413/+oo
,9448384o68o2779/+oo

,1623434489o9491/+oo	,3376565511905o3/+oo	,3376565511905o3/+oo	,1623434489o9491/+oo
,3oo4382o9361899/-o2	,2547668o2367929/-o2	-,2547668o2367929/-o2	-,3oo4382o9361899/-o2
,1124678195658?o/-o3	,1o1669456137282/-o2	,1o1669456137282/-o2	,1124678195658?o/-o3

,6237263798744443/-o1	-,11o8259722756623/-o1	,5215883267457383/-o2	-,134433o83o1155o5/-o2
,1677381428572613/+oo	,1613576121478633/+oo	-,1o81755o6159353/-o1	,229250209136777/-o2
,16oo5o946718123/+oo	,3494741o18o6443/+oo	,1762989o9o42644/+oo	-,5394694o4776988/-o2
,1636877796396o6/+oo	,33244o667923o51/+oo	,3487391494187o74/+oo	,9997o81o822o474/-o1

-,14o9726459456513/-o2	,7o2256884227946/-o3	-,8o4o67656235976/-o3	,2314o6485o94629/-o3
,6872794956102693/-o2	-,1717719929o6366/-o1	,186759951oo4463/-o2	-,388637998998192/-o3
,562o1261832397?/-o2	,69629355574o32o/-o2	-,2227253533799493/-o1	,864o2o76886472?/-o3
,624o17o67233256/-o2	,4291268391122623/-o2	-,4393o7916313o57/-o2	-,741849o64669453/-o2

,3653789o96?o1453/-o3	-,7747666530597?2/-o3	,3594o5624432543/-o3	-,1689o5732894293/-o3
,733o99423578953/-o3	,31o534o988o9661/-o2	-,6815660711385563/-o3	,2787218754523473/-o3
,6469317293499933/-o3	,6781733439375483/-o2	,2997826380140323/-o2	-,5828250613363o1/-o3
,6916974906846493/-o3	,574o7617438o443/-o2	,6874934o2129687/-o2	,3o942882o593771/-o3

V.3. Explicit Process of Orders m+s ($s \leq 5$)

In this section we shall give some explicit process with multiple nodes. General theorems on their existence, as with implicit methods, are not known, of course. One has rather to content one self with a laborious search for special solutions. The question for " optimal " methods (few nodes, small error) is still more difficult.

Thus we assume

(3.1 a) $a_1 = o$

(3.1b) $b_{ij}^{(k)} = o \quad (j=i,i+1.,,,.n \; ; \; k=1,...,m)$.

We further content ourselfes with methods which satisfy the following conditions:

1.) the first node ($a_1=o$) has multiplicity m ($m \geq 2$) ;

2.) all further nodes have multiplicity ≤ 2 ;
thus (cf. (1.18)):

(3.1c) $c_i^{(k)} = o \quad (i=2,...,n \; ; \; k=3,4,...,m)$

(3.1d) $b_{ij}^{(k)} = o \quad (j=2,...,i-1 \; ; \; k=3,...,m)$

3.) m and s satisfy the condition

(3.2) $m+s \leq 2m+2 \iff s \leq m+2$

Conditions for the Coefficients

Satisfaction of C(m) :

$$C(m) \Longleftrightarrow \sum_{\sigma=1}^{m} \frac{(k-1;-1;\sigma-1)}{\sigma!} \sum_{j=1}^{i-1} b_{ij}^{(\sigma)} a_j^{k-\sigma} \quad (3.1a) \quad \sum_{\sigma=1}^{k} \frac{(k-1;-1;\sigma-1)}{\sigma!} \sum_{j=2}^{i-1} b_{ij}^{(\sigma)} a_j^{k-\sigma} +$$

$$+\frac{1}{k}\,b_{i1}^{(k)} = \frac{a_i^k}{k} \qquad (i=1,\ldots,n\ ;\ k=1,\ldots,m)\ .$$

For i=1 and k=1,...,m this is satisfied because of $a_1=0$;
by putting

$$(3.3a)\qquad b_{21}^{(k)} = a_2^k \qquad (k=1,\ldots,m)$$

$$(3.3b)\qquad b_{ij}^{(k)} = a_i^k - \sum_{\sigma=1}^{k} \frac{(k;-1;\sigma)}{\sigma!} \sum_{j=2}^{i-1} b_{ij}^{(\sigma)} a_j^{k-\sigma} \qquad (i=3,\ldots,n\ ;\ k=1,\ldots,m)$$

C(m) is satisfied completely and the coefficients $a_2,\ldots,a_n$,
$b_{ij}^{(\sigma)}$ $(i=3,\ldots,n\ ;\ j=2,\ldots,i-1\ ,\ \sigma=1,2\ )$ are still free.

Satisfaction of B(m) :

$$B(m) \Longleftrightarrow \sum_{\sigma=1}^{m} (k-1;-1;\sigma-1) \sum_{i=1}^{n} c_i^{(\sigma)} a_i^{k-\sigma} \qquad (3.1a)$$

$$\sum_{\sigma=1}^{k} (k-1;-1;\sigma-1) \sum_{i=2}^{m} c_i^{(\sigma)} a_i^{k-\sigma} + (k-1)!\ c_1^{(k)} = \frac{1}{k} \qquad (k=1,\ldots,m)\ ;$$

by putting

$$(3.4)\qquad c_1^{(k)} = \frac{1}{k!} - \frac{1}{(k-1)!} \sum_{\sigma=1}^{k} (k-1;-1;\sigma-1) \sum_{i=2}^{n} c_i^{(\sigma)} a_i^{k-\sigma} \qquad (k=1,\ldots,m)$$

B(m) is satisfied and the coefficients $c_i^{(k)}$ $(i=2,\ldots,n\ ;\ k=1,2\ )$ are still free.

Using (3.2) the elementary weights can be distingished as follows:

1.) $\phi = [\phi_1 \ldots \phi_t]$ with $r_i \leq m$ $(i=1,\ldots,t)$:

because of C(m) Lemma 1 (from section V.2.) is applicable and all these weights can be reduced to $[\phi^{r-1}]$ (by (2.12)).

2.) $\phi = [\phi_1 \ldots \phi_t]$ with $r_1 > m$, $r_i \leq m$ $(i=2,\ldots,t)$:

all these can by Lemma 2 (cf. (2.16)) be reduced to $[\phi_1 \phi^{r-r_1-1}]$.

The condition (3.2) guarantees that all elementary weights of order $\leq m+s$ occur in the above cases (cf. p 101 , 2.))

We thus can restrict us to the elementary weights of the forms $[\phi^{r-1}]$ and $[\phi_1\phi^k]$ with $r \leq m+s$, $r_1 > m$, $k \geq 0$.

Because B(m) and the Corollary of Lemma 1 (in V.2.) by (3.3a,b) and (3.4) already all conditions for the orders $\leq m$ are satisfied. We thus can restrict us to the rest and determine the still free coefficients $a_2,\ldots,a_n$; $c_i^{(k)}$, $b_{ij}^{(k)}$ $(i=2,\ldots,n$; $j=2,\ldots,i-1$; $k=1,2)$ to satisfy the conditions for the orders $m+1,\ldots,m+s$.

We thus obtain the following conditions for the coefficients:

order m+1:

$$[\phi^m] = \frac{1}{m+1} \quad ;$$

order m+2 :

$$[\phi^{m+1}] = \frac{1}{m+2}$$

$$[[\phi^m]] = \frac{1}{(m+1)(m+2)}$$

order m+3 :

$$[\phi^{m+2}] = \frac{1}{m+3} \qquad [[[\phi^m]]] = \frac{1}{(m+1)(m+2)(m+3)}$$

$$[[\phi^{m+1}]] = \frac{1}{(m+2)(m+3)} \qquad [[\phi^m]\phi] = \frac{1}{(m+1)(m+3)}$$

Generally the additional conditions for order m+s ($s \leq m+2$;cf. (3.2)) are as follows :

$$[\ldots[[\phi^{m+k_1}]\phi^{k_2}]\ldots\phi^{k_\tau}] = \frac{1}{(m+k_1+1)(m+k_1+k_2+2)\ldots(m+k_1+\ldots+k_\tau+\tau)}$$

with

$$k_1+k_2+\ldots+k_\tau+\tau=s \quad \text{where } k_i \geq 0 \text{ and } \tau=1,\ldots,s \ .$$

As can be seen easly (induction !) for order m+s there exist 2^{s-1} conditions of this form.

Up to m+5 these are listed in /25/,p.45 .

The method of Fehlberg as special case

a) Fehlberg's method which,is known for its accuracy and its low expenditive of work, works as follows (cf. /12/,/13/):

diff. equation: $y' = f(x,y)$. $y(x_o) = y_o$

approximate solution of order m: $\hat{y}(x) = \sum_{k=o}^{m} \frac{h^k}{k!}\left[D^k y\right]_o = \sum_{k=o}^{m} h^k Y_k$

diff. equation for the approx. solution: $\hat{y}'(x) = \hat{f}(x) = \sum_{k=1}^{m} k h^{k-1} Y_k$

We put: $z(x) = y(x) - \hat{y}(x)$, where $y(x)$ is the exact solution and obtain for $z(x)$ the following diff. equation :

(3.9a) $z'(x) = y'(x) - \hat{y}'(x) = f(x,\hat{y}(x)+z(x)) - \hat{f}(x) =: \bar{f}(x,z(x))$

with $z(x_o) = z'(x_o) = \ldots = z^{(m)}(x_o) = o$.

This we now solve by a Runge-Kutta-process

(3.9b) $\hat{z}(x) = h \sum_{i=2}^{n} c_i k_i$

(3.9c) $k_i = f(x_o + a_i h, z_o + h \sum_{j=2}^{i-1} b_{ij} k_j)$ $(i=2,\ldots,n)$,

for which, because of the above made transformation, considerably higher orders are possible:
improved solution of order m+s:

(3.9d) $y_1(x) = \hat{y}(x) + \hat{z}(x) = \hat{y}(x) + h \sum_{i=2}^{n} c_i k_i$.

Remarks:

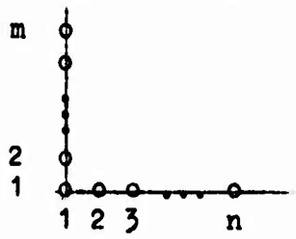
1. In /13/ Fehlberg gives a method of order m+4 using six nodes and with an astimation of the leading error term.
2. In /51/,p.101 Wanner has extended all the theory of Butcher to Fehlberg-processes.

b) Runge-Kutta-processes with one m-fold node:

Here we are interesting in process with one m-fold node and a fe

say n-1, single nodes .

Let this fact be expressed by the following diagramm :



The nodes of the method are pictured on the abscissa and the ordinates show their multiplicity. From this diagramm it is now clear why in (3.9b) and (3.9c) the indices i,j start with 2.

Both methods are now shown to be identical:

we have from (1.15,16) :

$$(3.10a)\quad y_2(x) = y_0 + h\sum_{i=1}^{n} c_i^{(1)} + h^2 c_1^{(2)} g_1^{(2)} + \ldots + h^m c_1^{(m)} g_1^{(m)}$$

with

$$(3.10b)\quad g_1^{(k)} = \left[D^k y\right]_0 = k!\,Y_k$$

$$(3.10c)\quad g_i^{(1)} = f\left(x_0 + a_i h,\ y_0 + h\sum_{j=1}^{i-1} b_{ij}^{(1)} g_j^{(1)} + \frac{h^2}{2!} b_{i1}^{(2)} g_1^{(2)} + \ldots + \frac{h^m}{m!} b_{i1}^{(m)} g_1^{(m)}\right)$$

$$(i=2,\ldots,n)$$

and it holds the following theorem:

Theorem: The method determined by (3.10a,b,c) is identical with Fehlberg's process (3.9a-d) if we put

$$(3.10d)\quad c_i^{(1)} = c_i \quad (i=2,\ldots,n)$$

$$(3.10e)\quad c_1^{(k)} = \frac{1}{(k-1)!}\left(\frac{1}{k} - \sum_{i=2}^{n} c_i^{(1)} a_i^{k-1}\right) \quad (k=1,\ldots,m)$$

$$(3.10f)\quad b_{ij}^{(1)} = b_{ij} \quad (i=2,\ldots,n\ ;\ j=2,\ldots,i-1)$$

$$(3.10g)\quad b_{i1}^{(k)} = a_i^k - \sum_{j=2}^{i-1} b_{ij}^{(1)} a_j^{k-1} \quad (i=2,\ldots,n\ ;\ k=1,\ldots,m) \quad .$$

Proof: We have to show that $y_1(x) = y_2(x)$:

Inserting (3.10b,c,e) into (3.10a) we obtain

$$y_2(x) = y_o+hY_1+\ldots+h^mY_m-\sum_{k=1}^{m}kh^k\sum_{i=2}^{n}c_i^{(1)}a_i^{k-1}Y_k+h\sum_{i=2}^{n}c_i^{(1)}g_i^{(1)} =$$

$$\hat{y}(x)+h\sum_{i=2}^{n}c_i^{(1)}\left(g_i^{(1)}-\sum_{k=1}^{m}k(a_ih)^kY_k\right)=$$

$$\hat{y}(x)+h\sum_{i=2}^{n}c_i^{(1)}\left(g_i^{(1)}-\hat{f}(x_o+a_ih)\right) =$$

$$\hat{y}(x)+h\sum_{i=2}^{n}c_i^{(1)}\bar{k}_i \qquad (3.11a)$$

with

$$\bar{k}_i = g_i^{(1)}-\hat{f}(x_o+a_ih) \quad (i=2,\ldots,n) \qquad (3.11b)$$

Next we insert (3.10b,f,g) into (3.10c):

$$g_i^{(1)}= f(x_o+a_ih,y_o+hb_{i1}^{(1)}g_1^{(1)}+\ldots+\frac{h^m}{m!}b_{i1}^{(m)}g_1^{(m)}+h\sum_{j=2}^{n}b_{ij}^{(1)}g_j^{(1)}) =$$

$$f(x_o+a_ih,y_o+a_ihY_1+\ldots+(a_ih)^mY_m-\sum_{k=1}^{m}kh^k\sum_{j=2}^{i-1}b_{ij}^{(1)}a_j^{k-1}Y_k+h\sum_{j=2}^{n}b_{ij}^{(1)}g_j^{(1)})$$

$$f(x_o+a_ih,\hat{y}(x_o+a_ih)+h\sum_{j=2}^{i-1}b_{ij}^{(1)}(g_j^{(1)}-\sum_{k=1}^{m}k(a_jh)^{k-1}Y_k)) \qquad (3.11b)$$

$$f(x_o+a_ih,\hat{y}(x_o+a_ih)+h\sum_{j=2}^{i-1}b_{ij}^{(1)}\bar{k}_j) \quad (i=2,\ldots,n) \qquad (3.11c)$$

From (3.11b,c) we have

$$\bar{k}_i = f(x_o+a_ih,\hat{y}(x_o+a_ih)+h\sum_{j=2}^{i-1}b_{ij}^{(1)}\bar{k}_j) -\hat{f}(x_o+a_ih) \quad (i=2,\ldots,n) \quad ; \qquad (3.12a)$$

On the other side it follows from (3.9a,c) :

$$k_i = f(x_o+a_ih,\hat{y}(x_o+a_ih)+h\sum_{j=2}^{i-1}b_{ij}k_j)-\hat{f}(x_o+a_ih) \quad (i=2,\ldots,n) \; . \qquad (3.12b)$$

Next compare the conditions which are to be satisfied by $c_i^{(1)}$ and $b_{ij}^{(1)}$ $(i=2,\ldots,n \; ; \; j=2,\ldots,i-1)$ with those for the method of Fehlberg (e.g. in the form given by Wanner /51/,p.103). We thereby confirm easily that the equations coincide; thus we may put

$b_{ij}^{(1)}=b_{ij}$ and $c_i^{(1)}=c_i$ ($i=2,\ldots,n \; ; \; j=2,\ldots,i-1$) .

Hence from (3.12a,b) we have $k_i=\bar{k}_i$ (i=2,...,n) and (3.11a),(3.9d) shows that $y_1(x)=y_2(x)$.

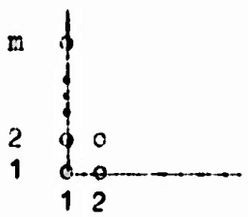
The coefficients (3.10e,g) are of course determined by (3.3),(3.4) .
Done.

Some explicit methods of orders (m+2),...,(m+5)

Here we list some explicit formulas of different orders. Their derivation from the above conditions is given in full detail in the thesis /25/, pp 51-61 .

1.) Formula of order m+2 :

diagram :



n=2
m ⩾ 2
order m+2

Coefficients:

$a_1=0$, $a_2=\frac{m}{m+2}$;

$c_1^{(1)}=1$, $c_2^{(1)}=0$, $c_2^{(2)}=\frac{1}{a_2^m(m+1)(m+2)}$;

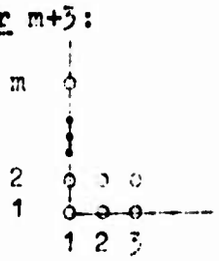
$c_1^{(k)}=\frac{1}{(k-1)!}\left(\frac{1}{k}-(k-1)c_2^{(2)}a_2^{k-2}\right)$ (k=2,...,m)

$b_{21}^{(k)}=a_2^k$ (k=1,...,n) .

No coefficient can be choosed freely.

2.) Formula of order m+3:

diagram



n=2
m ⩾ 2
order m+3

coefficients:

$a_1=0$, $a_2=\frac{m}{m+3}$, $a_3=1$;

$c_1^{(1)}=1$, $c_2^{(1)}=c_3^{(1)}=0$;

$c_2^{(2)}=\frac{2}{3(m+1)(m+2)a_2^m}$, $c_3^{(2)}=\frac{1}{3(m+1)(m+2)}$;

$c_1^{(k)}=\frac{1}{(k-1)!}\left(\frac{1}{k}-(k-1)(c_2^{(2)}a_2^{k-2}+c_3^{(2)})\right)$ $(k=2,\dots,m)$;

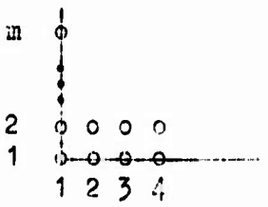
$b_{21}^{(k)}=a_2^k$ $(k=1,\dots,m)$;

$b_{32}^{(1)}=0$, $b_{32}^{(2)}=\frac{6}{a_2^{m-1}m(m+3)}$;

$b_{31}^{(k)}=1-\frac{k(k-1)}{2}b_{32}^{(2)}a_2^{k-2}$ $(k=1,\dots,m)$.

3.) Formula of order m+4 :

diagram:



n=4
m ≥ 2
order m+4 .

Coefficients :

$a_1=0$, $a_2=\frac{m}{m+6}$, $a_3=\frac{m+2}{m+4}$, $a_4=1$;

$c_1^{(1)}=1$, $c_2^{(1)}=c_3^{(1)}=c_4^{(1)}=0$;

$c_2^{(2)}=\frac{(m+6)^2}{12a_2^m(m+1)(m+2)(m+3)^2}$, $c_3^{(2)}=\frac{3(m+4)}{4a_3^m(m+1)(m+3)^2}$,

$c_4^{(2)}=\frac{m}{6(m+1)(m+2)(m+3)}$;

$c_1^{(k)}=\frac{1}{(k-1)!}\left(\frac{1}{k}-(k-1)(c_2^{(2)}a_2^{k-2}+c_3^{(2)}a_3^{k-2}+c_4^{(2)}a_4^{k-2})\right)$ $(k=2,\dots,m)$

$b_{21}^{(k)}=a_2^k$ $(k=1,\dots,m)$

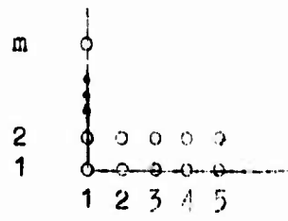
$$b_{32}^{(1)}=0\ ,\ b_{32}^{(2)}=\frac{2}{c_3^{(2)}a_2^m(m+1)(m+2)(m+3)(m+6)}\ ;$$

$$b_{42}^{(2)}=-\frac{1}{c_4^{(2)}a_2^{m-1}(m+1)(m+2)(m+3)^2}\ ,\ b_{43}^{(2)}=\frac{1}{c_4^{(2)}a_3^{m-1}(m+1)(m+2)(m+3)^2}$$

$$b_{i1}^{(k)}=a_i^k-\frac{k(k-1)}{2}\sum_{j=2}^{i-1}b_{ij}^{(2)}a_j^{k-2}\quad(i=3,4\ ;\ k=2,\dots,m)\ .$$

4.) Formula of order m+5 :

diagram:



n=5
m ≥ 3
order m+5 .

Coefficients:

$$a_1=0\ ,\ a_2=\frac{m}{m+5}\ ,\ a_3=\frac{m}{m+3}\ ,\ a_4=\frac{m+2}{m+5}\ ,\ a_5=1\ ;$$

$$c_1^{(1)}=1\ ,\ c_2^{(1)}=c_3^{(1)}=c_4^{(1)}=c_5^{(1)}=0\ ;$$

$$c_2^{(2)}=\frac{3m^2}{10a_2^{m+2}(m+1)(m+2)(m+3)(m+4)}\ ,\ c_3^{(2)}=-\frac{m^2}{6a_3^{m+2}(m+1)(m+2)(m+4)}\ ,$$

$$c_4^{(2)}=\frac{(m+2)^2}{6a_4^{m+2}(m+1)(m+3)(m+4)}\ ,\ c_5^{(2)}=\frac{3m^2+15m+10}{15(m+1)(m+2)(m+3)(m+4)}\ ;$$

$$c_1^{(k)}=\frac{1}{(k-1)!}\left(\frac{1}{k}-(k-1)(c_2^{(2)}a_2^{k-2}+\dots+c_5^{(2)}a_5^{k-2})\right)\quad(k=2,\dots,m)\ ;$$

$$b_{21}^{(k)}=a_2^k\quad(k=1,\dots,m)$$

$$b_{32}^{(1)}=b_{32}^{(2)}=0\ ;$$

$$b_{42}^{(1)}=b_{43}^{(1)}=0\ ,\ b_{42}^{(2)}=\frac{4}{3c_4^{(2)}a_2^m(m+1)(m+2)(m+3)(m+4)}\ ,\ b_{43}^{(2)}=0\ ;$$

$$b_{52}^{(1)}=b_{53}^{(1)}=b_{54}^{(1)}=0\ ,$$

$$b_{52}^{(2)}=\frac{7}{6c_5^{(2)}a_2^m(m+1)(m+2)(m+3)(m+4)}\ ,\ b_{53}^{(2)}=-\frac{m+3}{2c_5^{(2)}a_3^{m-1}m(m+1)(m+2)(m+4)}$$

$$b_{54}^{(2)} = \frac{1}{2c_5^{(2)} a_4^m (m+1)(m+3)(m+4)} \quad ;$$

$$b_{i1}^{(k)} = a_i^k - \frac{k(k-1)}{2} \sum_{j=2}^{i-1} b_{ij}^{(2)} a_j^{k-2} \qquad (k=1,\ldots,m \; ; \; i=3,4,5 \;)$$

Numerical Examples

The practical evaluation of these formulas is only valuable, when it is combined with the use of the recursion formulas which are described in Chapter II. With these the methods again can fully be made automatic by using the same subroutines. It may finally be noted that in many cases the calculation of Df, D^2f,... often requires much less work than the calculation of f itself (especial if there occur a lot of elementary functions like exp, log, sin, cos, ...).

In the following the methods are tested at some differential equations with known solution. They are further compared with the method of Fehlberg and with the power-series method. All computations are with order 10 and have been carried out with single precisions (9D) on the Zuse Z23 computer.

Example 1 : $y' = 2x(e^{-x^2} - y) \quad y(0)=1$;

solution: $y(x) = (1+x^2)e^{-x^2}$.

Example 2 : $y' = \frac{1}{2}y^2 x^{-3/2} \quad y(1)=1$;

solution: $y(x) = \sqrt{x}$.

Example 3 : $y' = 1 - e^{-y}(\sin x - \cos x) \quad y(0)=0$;

solution : $y(x) = \log(\sin x + e^x)$.

Example 4 : $y' = \cos x \,.(y+\sin x) \quad y(0)=1$

solution: $y(x) = 2e^{\sin x} - \sin x - 1$.

Example 5 : $y' = (x^4+y^4)/xy^3$ $y(1)=1$,

solution: $y(x) = x.(1+4.\log x)^{1/4}$

Example 6 : $y' = 2(xy^{3/2}-y)$ $y(0)= 0,25$,

solution: $y(x) = (e^x+x+1)^{-2}$.

Example 7 : $y' = (xy^2+y)/(x.\log x)$ $y(e)= 0,5$

solution: $y(x) = \log x/(e+2-x)$.

In the following table the errors of the different methods with these examples and with the given step sizes are listed.

Example	h	power series	I	II	III	IV	Fehlberg
1.	0,5	$1,6.10^{-6}$	$1,6.10^{-7}$	$4,3.10^{-8}$	$1,5.10^{-8}$	$8,2.10^{-8}$	$2,2.10^{-7}$
2.	0,8	$4,0.10^{-4}$	$3,4.10^{-6}$	$8,1.10^{-7}$	$4,4.10^{-8}$	$7,6.10^{-8}$	$6,4.10^{-7}$
3.	0,25	$5,2.10^{-6}$	$1,0.10^{-7}$	$4,4.10^{-8}$	$8,5.10^{-10}$	$1,0.10^{-8}$	$5,0.10^{-8}$
4.	1	$4,9.10^{-4}$	$3,4.10^{-6}$	$2,8.10^{-6}$	$3,2.10^{-6}$	$1,0.10^{-6}$	$1,3.10^{-6}$
5.	0,2	$1,4.10^{-3}$	$1,5.10^{-5}$	$4,2.10^{-6}$	$3,7.10^{-7}$	$4,8.10^{-7}$	$1,1.10^{-6}$
6.	0,4	$9,2.10^{-6}$	$5,2.10^{-7}$	$1,9.10^{-7}$	$2,4.10^{-8}$	$4,3.10^{-8}$	$1.0,1.10^{-7}$
7.	0,7	$1,1.10^{-5}$	$1,8.10^{-6}$	$1,0.10^{-6}$	$7,7.10^{-8}$	$4,2.10^{-7}$	$1,3.10^{-6}$

h= step size
I : formula of order m+2 ;
II : formula of order m+3 ;
III : formula of order m+4
IV : formula of order m+5 .

As can be seen, the results of formulas III and IV on the average have the same accuracy than the method of Fehlberg. It can further be seen, that with equal order the methods with more nodes are very much better. The results of III, IV and Fehlberg are mostly 2-3 digits better than with the power series method of the same order. In addition note that

the necessary work for the power series method mostly is higher than with the other formulas.

The following questions are still open:

1. " optimal" methods: the coefficients in general are not uniquely determined by the conditions. Some of them have been fixed arbitrary, mostly to reach simple results.
   How are they to be fixed to give methods with minimal error?
2. How are effective error estimates possible ?
3. How is the stability of the methods ?

Chapter VI

On Step-size Control

by G. Wanner

This chapter deals with the problem of choosing the step sizes in the numerical integration of ordinary differential equations using one-step methods. First the frequently used formulas are discussed which try to keep the local error constant. Then expressions for an "optimal" step-size control are developed which take into account the propagation of the local errors to the final result. Numerical results are given and compared with those of the step-size control of Merrison.

VI.1 Step-size Control

A system of n ordinary differential equations

$$y' = f(x,y)$$

is given and the solution $y(x)$ with initial values x_o, y_o is wanted at some point x_N. Using some one-step method, the integration proceeds on the steps $x_o<x_1<x_2<\ldots<x_N$ with the step-sizes $h_1=x_1-x_o$, $h_2=x_2-x_1,\ldots,h_N=x_N-x_{N-1}$. For a step size control the method has to be equipped with some error estimation, i.e., to each initial point x_k, y_k and step-size h_{k+1} it gives a approximation $\hat{y}_i(x_{k+1})$ to the solutions and approximate error estimations R_i. The usual procedure now is trying to keep these local errors equal to some given numbers γ_i, the wanted errors. These might be $10^{-5}, 10^{-10}, 10^{-20}$ and depend on the wanted accuracy. Thus by putting

$$\eta = \max_i \frac{|R_i|}{\gamma_i} \tag{1.1}$$

one tries to keep $\eta=1$. A possible procedure is now the following: The first step is calculated with a guessed step-size h_1. Then η can be evaluated by (1.1). Of course, η will not be equal 1. If p is the order of the method, a much more better step-size would have been

$$\bar{h} = h_1 \sqrt[p+1]{1/\eta}. \tag{1.2}$$

But if η is not very much greater than 1, say $\eta<\eta_2$ with $\eta_2 = 1.5$ or 10, then we use $\bar{h}$ for the next step

$$h_2=\bar{h},$$

otherwise we repeat the first step with the step size $\bar{h}\to h_1$. The same procedure is then also used in the following steps $h_2, h_3, \ldots$.

VI.2. Damping

Occasionally, especially in regions where R_i changes sign, η may be very small or even zero. In such cases, formula (1.2) would lead to an excessive increase of the step size. For this

reason, one chooses a number $\eta_1<1$ (say $=1/10$, $1/100$) and if $\eta<\eta_1$, one replaces (1.2) by

$$(2.1) \qquad \bar{h} = h_1 \frac{p+2}{p+1} \left(1 - \frac{\eta}{(p+2)\,\eta_1}\right) \sqrt[p+1]{1/\eta_1} \qquad (\text{if } \eta<\eta_1).$$

Thus the step size can increase at most by the factor

$$\frac{p+2}{p+1} \sqrt[p+1]{1/\eta_1}\,.$$

This stabilizes the step size control and guards against overflow. Formula (2.1) is obtained by replacing the hyperbola $\sqrt[p+1]{1/\eta}$ by its tangent of the point η_1.

VI.3. Morrison's Control

Consider the example

$$(3.1) \qquad y' = y^2,\ y(0)=1,\ y(0.999999)=?$$

The solution is $y=1/(1-x)$, $y(0.999999)=10^6$. The solution for a general initial value y_0 is $y(x,y_0) = 1/(1/y_0-x)$. The derivative of this solution with respect to that initial value is

$$(3.2) \qquad H(x) = \frac{\partial y(x,y_0)}{\partial y_0} = 1/(1/y_0-x)^2 y_0^2 = y^2/y_0^2.$$

Thus, if the initial value $y_0=1$ is changed, say, by 10^{-15}, then the solution at the point 0.999999 changes by 10^{-3} since $H(0.999999)=10^{12}$. If we compute this example by a stepwise numerical integration with local accuracy 10^{-15}, the final result will not be better than 10^{-3}. Of course here it is unwise to compute also the last steps with this same accuracy. The last steps need not to be calculated with the same accuracy than the first. The idea lies at hand to multiply the chosen error sizes γ_i by the connection matrix $H(x)$ along the solution. This means to replace (1) by

(3.3) $$\eta = \max_i \left| \frac{R_i}{(H(x)Y)_i} \right| .$$

This is the step size control, which Morrison /37/ has proved to be nearly "optimal" for the case when n=1 (one equation only) and when the errors R_i all have the same sign.

VI.4. Another Possibility

There is still another possibility for a step size control which shall be derived now:
Assume the differential equation to be integrated from x_o to x_N using N steps $h_1=x_1-x_o$, $h_2=x_2-x_1$,... . The local error of the j-th step we denote by e_j and its propagation to the final result is

(4.1) $$e_j^{(N)} = H(x_N)H^{-1}(x_j)e_j .$$

We again assume that n=1 and that all errors have constant sign, although the results may be interpreted for the other cases as well.
Neglecting rounding errors we may assume $e_j=\phi_j h_j^{p+1}$, thus $e_j^{(N)} = \chi_j h_j^{p+1}$ where $\chi_j=H(x_N)H^{-1}(x_j)\phi_j$.
Assuming that χ_j does not depend on $h_1,..,h_j$ (what however actually is the case), we solve the easy minimum problem

$$\sum_{j=1}^{N} e_j^{(N)} = \sum_{j=1}^{N} \chi_j h_j^{p+1} = \min!$$

under the condition that

$$\sum_{i=1}^{N} h_j = x_N - x_o = C_o .$$

The method of Lagrange multipliers gives $h_j = C_1 / \sqrt[p]{\chi_j}$.
Thus the local error e_j should be

$$e_j=\phi_j h_j^{p+1}=\frac{\phi_j C_2}{x_j \sqrt[p]{x_j}}=\frac{k\, h_j}{H(x_N)H^{-1}(x_j)}$$

Hence, the step-sizes are chosen "optimal", if

(4.2) $H(x_N)H^{-1}(x_j)e_j h_j^{-1} = k,$

i.e., if the contribution $e_j^{(N)}$ of each step to the final result is proportional to the step-size.

In the case n=1 $H(x_N)$ is only a constant number and need not be known. Hence, in the course of computation, it is only necessary to keep

(4.3) $H^{-1}(x_j)e_j h_j^{-1}=\gamma.$

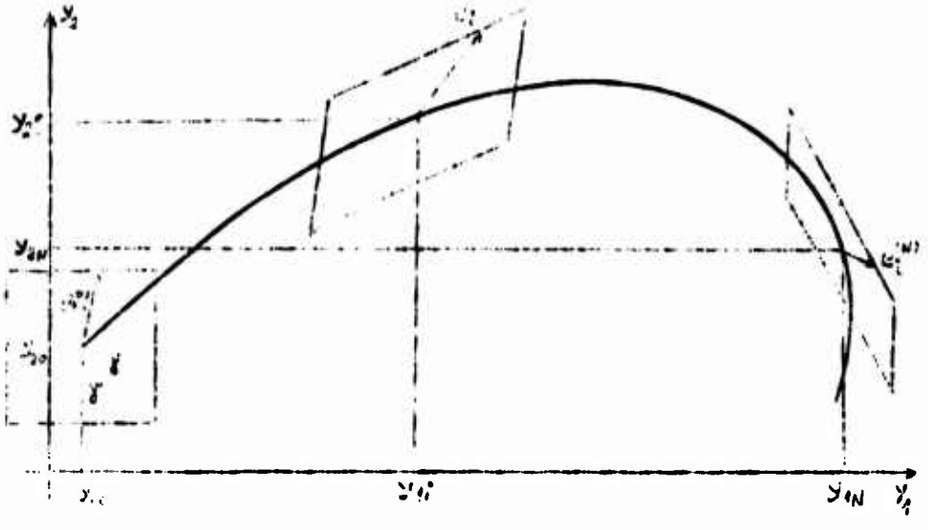
This result differs from that of Morrison by the denominator h_j. An error estimation for the total truncation error is now obtained as follows:

The final error $e_j^{(N)}$ (4.1), which results from e_j is because of (4.3) equal to $\pm h_j H(x_N)\gamma$. These numbers are summed up easily to give

(4.4) $E = (x_N-x_o)H(x_N)\gamma$

as estimation of the total truncation error of the final result.

Formula (4.2) may also be interpreted for systems of differential -quations. But then the knowledge of the final connectionmatrix $H(x_N)$ is necessary. On the other hand, the use of (4,3)seems only adequate, of all eigenvalues of $H(x_N)$ have approximately the same size. This is, because (4.3) mapps the error not to the endpoint x_N, but to the initial point x_o and the box $\max_j|R_j|=\gamma$ may change shape considerably



Finally we mention the paper Greenspan-Hafner-Ribaric /20/. There Morrison's control is compared with several others (such as the "natural" step size of Collatz /7/,p.89). For differential equations with constant coefficients (y'=cy) Morrison's control as well as the above "optimal" control become a control with constant step size (for the methods considered in this report).

VI.5. Numerical Examples

Using the Lie-series method of Chapter III, the different controls of the sections VI.1, VI.3 and VI.4 have been compared at several examples. The results are listed below:

Example 1: $y' = y^2$, $y(0) = 1$, solution $y(x) = 1/(1-x)$

step size control	m	s	k	γ	x	actual error of $\bar{y}(x)$	error estimation for $\bar{y}(x)$	steps	time per step (m sec)
normal	10	3	2	10^{-17}	0.9	$3.01*10^{-15}$	---	23	44
					0.99	$3.05*10^{-13}$	---	50	
					0.999999	$2.89*10^{-5}$	---	210	
	15	3	2	10^{-19}	0.9	$1.65*10^{-17}$	---	15	60
					0.99	$1.67*10^{-15}$	---	32	
					0.999999	$1.73*10^{-7}$	---	123	
optimal 1	10	3	2	10^{-17}	0.9	$6.8*10^{-16}$	$9.0*10^{-16}$	25	56
					0.99	$7.5*10^{-14}$	$9.9*10^{-14}$	50	
					0.999999	$7.6*10^{-6}$	$10.0*10^{-6}$	148	
	15	3	2	10^{-19}	0.9	$6.4*10^{-18}$	$9.0*10^{-18}$	16	71
					0.99	$7.0*10^{-16}$	$9.9*10^{-16}$	31	
					0.999999	$7.2*10^{-8}$	$10.0*10^{-8}$	92	

Up to 0.99 only, the optimal step size control gives no noticeable increase of effectiveness.

Example 2: $y_1'=y_2$, $y_2'=y_1$, $y_1(0)=0$, $y_2(0)=1$, solutions: $y_1 = \sinh x$, $y_2 = \cosh x$

Results for x=10:

step size control	m	s	k	γ	h	actual error of $y_i(10)$	error estimat. for y_i	steps
normal	13	5	3	10^{-20}	0.83 ..0.53	$1.21*10^{-16}$	---	15
optimal	13	5	3	10^{-20}	0.805 ..0.811	$3.75*10^{-16}$	$4.4*10^{-16}$	13
optimal	13	5	3	10^{-21}	0.7203..0.7205	$4.2 *10^{-17}$	$4.9*10^{-17}$	14

As expected for linear systems with constant coefficients, the optimal step size remained constant. The estimate for the total propagated error is satisfactory.

Example 3: $y'=-xy^3$, $y(-1) = y_o$, solution $y(x) = y_o(1+(x^2-1)y_o^2)^{-\frac{1}{2}}$.

a) $y_o = 0.999999995$, thus $y(0) \doteq 10000$

$y(1) = 0.999999995$.

Results for m=15, s=3, k=2:

step size control	γ	h	x	actual error of $y(x)$	error estimat. for $y(x)$	steps
normal	10^{-10}	0.2 ... 0.00003	0	$1.9*10^{+1}$	----	28
			1	$1.8*10^{-11}$	----	67
optimal	10^{-12}	0.2 ... 0.00007	0	$4.9*10^{-1}$	$9.0*10^{-1}$	21
			1	$4.1*10^{-13}$	$20.0*10^{-13}$	55

For this example, constant step size is not advisable.

b) y_o = 0.99999999995, thus y(0) ≙ 100000
y(1) = 0.99999999995.

Results for m=15, s=3, k=2:

step size control	γ	h	x	actual error of y(x)	error estimat. for $\bar{y}$(x)	steps
normal	10^{-12}	0.29..0.0000023	0	$1.6\ 10^{+2}$	---	46
			1	$1.7\ 10^{-13}$	---	107
optimal	10^{-14}	0.24..0.0000060	0	$5.1\ 10^{-0}$	$9.6\ 10^{-0}$	32
			1	$5.4\ 10^{-15}$	$20.0\ 10^{-15}$	80

Chapter VII

Calculation of Switch-on Transients at the telegraphic Equation by Means of Generalized LIE - Series

by R. Saely

Abstract

This chapter deals with the switch-on transients occurring in the telegraphic equation,i. e., an initial and boundary value problem of a hyperbolic partial differential equation, by means of generalized Lie series. We shall assume that an ordinary alternating voltage $U(o,t) = A \cos \omega t + B \sin \omega t$ is applied across the input terminals of a telegraphic line (electric twin line) of length a. We confine our investigations to two limiting cases, namely, that the line is either shorted or open at the other end.

The first part of the paper gives a formal solution using power series. The solution is represented by means of Lie series with a generalized Lie - Operator. Next the switch-on transients is treated with shorted wires and given initial and boundary conditions. Two numerical examples shall illustrate this switch-on transients problem. Finally the computation of the solution $U(x,t)$ and $J(x,t)$ for the initial and boundary value problem with open wires is given.

My thank go to Prof. W.Groebner for his assistance;I wish acknowledge the discussions with H.Reitberger, G.Wanner and K.H. Kastlunger.

VII.1 Introduction

1. The Telegraphic Equation

In the present paper we shall calculate the switch-on transients occurring in the telegraphic equation, i. e., an initial and boundary value problem of a hyperbolic partial differential equation, by means of generalized Lie-series. We shall assume that an ordinary alternating voltage $U(o,t) = A \cos \omega t + B \sin \omega t$ is applied across the input terminals of a telegraphic line (electric twin line) of length a. We confine our investigations to two limiting cases, namely, that the line is either shorted or open at the other end.

Let a_1, a_2 be the input and e_1, e_2 the output terminals of the line. We take one axis of coordinates along the line and denote the distance from the input terminals by x. The length of the line is a. At time t, the current J(x,t) flows in the wire at the point x, the voltage between the two wires is U(x,t).

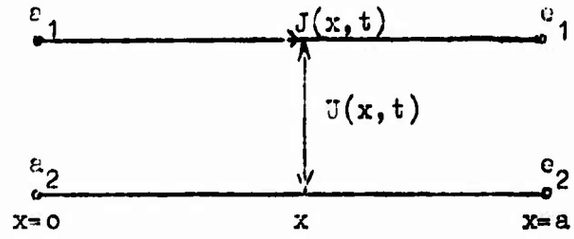


Fig. 1. Schematic diagram of a telegraph line

Of these parallel wires we consider a very small line element I,II,III,IV of length dx (infinitesimal four-pole). We assume the line constants, referred to unit length, to be independent of space and time coordinates and denote them by the following symbols:

r ..... resistance
l ..... inductive reactance
q ..... conductance (leaking insulation)
c ..... capacitive reactance

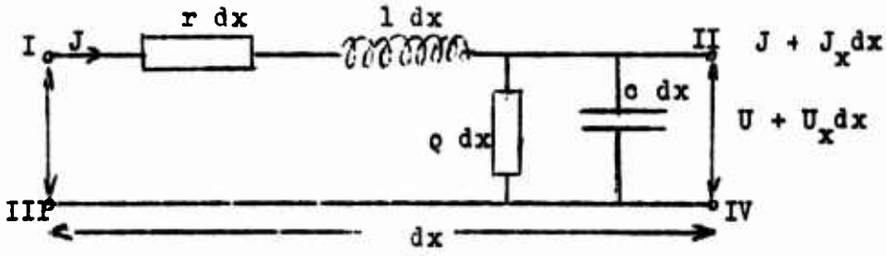


Fig. 2. Infinitesimal element of a telegraphic line

The two basic equations of the telegraphic equation follow from the laws of electromagnetic theory:

$$J_x(x,t) = -gU(x,t) - cU_t(x,t) \tag{1.1}$$

$$U_x(x,t) = -rJ(x,t) - lJ_t(x,t) \tag{1.2}$$

To eliminate current in (1.2) we differentiate the first equation with respect to t and the second with respect to x. Putting

$$rg = \alpha$$
$$rc + gl = \beta$$
$$lc = \gamma$$

we obtain the telegraphic equation for the voltage U(x,t):

$$U_{xx}(x,t) = \alpha U(x,t) + \beta U_t(x,t) + \gamma U_{tt}(x,t) \tag{1.3}$$

An analogous equation in J(x,t) can be found by differentiating (1.1) with respect to x and eliminating $U_x(x,t)$ by means of (1.2):

$$J_{xx}(x,t) = \alpha J(x,t) + \beta J_t(x,t) + \gamma J_{tt}(x,t) \tag{1.4}$$

In their physical meaning, the constants α, β and γ are positive. Mathematical treatment requires only that $\gamma > 0$, because this makes the equation hyperbolic. With $\alpha = \beta = 0$, the telegraphic equation becomes the ordinary wave equation.

2. Formal Solution of the Telegraphic Equation

Equation (1.3) describes all electric phenomena in the two parallel wires. A problem frequently arising is: what happens when the system is switched on? At given time (t=o) the system, whose electric condition at that moment is known, experiences some kind of external influence. We wish to calculate the changes produced by this influence. We assume this influence to be a suddenly applied voltage. We assume further that, at the point x=o, the voltage $U(o,t) = f(t)$ is a given function of time. From the basic equation (1.2) we find

$$U_x(o,t) = -rJ(o,t) - lJ_t(o,t) = - rg(t) - lg_t(t) = h(t)$$

Now we shall solve the telegraphic equation with the power series expansion (cf. /23/ , p.112)

$$U(x,t) = \sum_{\nu=0}^{\infty} x^{\nu} \Psi_{\nu}(t) \tag{1.5}$$

where the functions $\Psi_{\nu}(t)$ are yet to be determined. The functions $\Psi_0(t)$ and $\Psi_1(t)$ can be found from the initial conditions:

$$U(o,t) = f(t) = \Psi_0(t) \qquad \text{and} \tag{1.6}$$

$$U_x(o,t) = h(t) = \Psi_1(t) \tag{1.7}$$

The remaining functions $\Psi_{\nu}(t)$ may be calculated by means of a recursion formula which can be obtained by comparing coefficients of x^{ν} after the power series expansion has been inserted in (1.3).

$$\Psi_{\nu+2}(t) = \frac{\alpha\Psi_{\nu}(t) + \beta\Psi'_{\nu}(t) + \gamma\Psi''_{\nu}(t)}{(\nu+1)(\nu+2)} \tag{1.8}$$

All functions $\Psi_{\nu}(t)$ can be calculated from this formula, because $\Psi_0(t)$ and $\Psi_1(t)$ are known from the initial conditions (1.6) and (1.7).

Introducing the linear operator

$$D = \alpha + \beta\frac{\partial}{\partial t} + \gamma\frac{\partial^2}{\partial t^2} \tag{1.9}$$

we can write the solution of the differential equation in a more convenient form.

The recursion formula is then

$$\Psi_{\nu+2}(t) = \frac{1}{(\nu+1)(\nu+2)} D\Psi_\nu(t) \tag{1.8'}$$

The function (series) solving the equation can now be written as the sum of the terms $x^\nu \Psi_\nu$ with ν even ($\nu = 0,2,4,\ldots$) and of the terms with ν odd ($\nu = 1,3,5,\ldots$).

For the Ψ_ν with even index we write

$$\nu + 2 = 2\mu$$

This new summation index μ is inserted in the recursion formula (1.8')

$$\Psi_{\nu+2}(t) = \Psi_{2\mu} = \frac{1}{2\mu(2\mu-1)} D\Psi_{2\mu-2} = \frac{1}{(2\mu)!} D^\mu \Psi_0(t) = \frac{1}{(2\mu)!} D^\mu f(t)$$

Likewise, for odd ν we have:

$$\nu + 2 = 2\mu + 1$$

$$\Psi_{\nu+2}(t) = \Psi_{2\mu+1}(t) = \frac{1}{2\mu(2\mu+1)} D\Psi_{2\mu-1} = \frac{1}{(2\mu+1)!} D^\mu \Psi_1(t) = \frac{1}{(2\mu+1)!} D^\mu h(t)$$

Hence follows the formal solution of the telegraphic equation (1.5)

$$U(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^\nu f(t) + \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^\nu h(t) \tag{1.10}$$

To obtain the complete solution one has to know the corresponding boundary and initial conditions.

The current $J(x,t)$ is obtained by integration of the differential equation (1.1)

$$J(x,t) = J_0(t) - \int (\varrho + c\frac{\partial}{\partial t}) U(x,t)\, dx \tag{1.11}$$

The proof that (1.10) converges can be given by a majorant method; it can be found in /23/, p. 114 et sequ.).

VII.2 Switch-on Transients with Shorted Wires

1. Initial and boundary conditions

To be able to solve the telegraphic equation, i. e.,to describe the boundary and initial value problem completely, we need the corresponding boundary and initial conditions.

We shall assume the following conditions:

Until time t=o there is no current nor voltage in the wires. An alternating voltage $U(o,t) = A \cos \omega t + B \sin \omega t$ is applied at this moment t=o. This immediately gives the initial conditions for the voltage function U(x,t) at any x except at x=o and for the current function J(x,t).

(2.1) $$U(x,o) = o \qquad x > o$$

(2.2) $$J(x,o) = o$$

The initial condition

$$U_t(x,o) = o \qquad x > o$$

follows from Eq. (1.1).

What we still need are the boundary conditions for both ends of the line of length a. One of them, for x=o (after an ordinary alternating voltage has been applied), is

(2.3) $$U(o,t) = A \cos \omega t + B \sin \omega t = f(t)$$

As we assume the line to be shorted, the other boundary condition for x=a at the end of the line is

(2.4) $$U(a,t) = o$$

2. Transformation of a few expressions

Before introducing the initial and boundary conditions into the formal solution for $U(x,t)$ (cf.(1.1o)) we shall bring a few expressions into a more convenient form.

We apply the generalized Lie-operator D of Eq.(1.9) to the function $f(t)$ which gives the variation of the voltage $U(0,t)$ at the point $x=0$

$$D^0 f(t) = f(t) = A_0 \cos \omega t + B_0 \sin \omega t$$

$$D^1 f(t) = Df(t) = A_1 \cos \omega t + B_1 \sin \omega t =$$

$$= [(\alpha - \gamma\omega^2)A_0 + \beta\omega B_0] \cos \omega t + [-\beta\omega A_0 + (\alpha - \gamma\omega^2)B_0] \sin \omega t$$

This gives the following relations for the coefficients A_1 and B_1:

$$A_1 = (\alpha - \gamma\omega^2)A_0 + \beta\omega B_0$$

$$B_1 = (\alpha - \gamma\omega^2)B_0 - \beta\omega A_0$$

By applying the operator D ν-times to $f(t)$,

$$D^\nu f(t) = A_\nu \cos \omega t + B_\nu \sin \omega t \tag{2.5}$$

we obtain recursion formulas for the coefficients A_ν and B_ν (proof by induction):

$$A_\nu = (\alpha - \gamma\omega^2)A_{\nu-1} + \beta\omega B_{\nu-1} \tag{2.6}$$

$$B_\nu = (\alpha - \gamma\omega^2)B_{\nu-1} - \beta\omega A_{\nu-1} \tag{2.7}$$

for $\nu \geq 1$

With the matrix

$$\Omega = \begin{pmatrix} \alpha - \gamma\omega^2 & -\beta\omega \\ +\beta\omega & \alpha - \gamma\omega^2 \end{pmatrix} \tag{2.8}$$

and the corresponding transposed Ω^T (for the rules of matrix calculus cf. /24/) the above formulas can be written in matrix form:

$$\begin{pmatrix} A_\nu \\ B_\nu \end{pmatrix} = \Omega^T \begin{pmatrix} A_{\nu-1} \\ B_{\nu-1} \end{pmatrix} = (\Omega^T)^2 \begin{pmatrix} A_{\nu-2} \\ B_{\nu-2} \end{pmatrix} = (\Omega^T)^\nu \begin{pmatrix} A_0 \\ B_0 \end{pmatrix}$$

After having transposed this matrix equation,

$$(A_\nu, B_\nu) = (A_0, B_0)\left[(\Omega^T)^\nu\right]^T = (A_0, B_0)\,\Omega^\nu$$

we insert the matrix (A_ν, B_ν) in the first term of the formal solution for $U(x,t)$ (cf.(1.1o)):

$$\sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^\nu f(t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} (A_\nu \cos \omega t + B_\nu \sin \omega t) =$$

$$= \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} (A_\nu, B_\nu) \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix}$$

$$\sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^\nu f(t) = (A_0, B_0)\left(\sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} \Omega^\nu \right) \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} \tag{2.9}$$

3. Statement for $h(t)$

For the function $h(t) = U_x(o,t)$ (see (1.7)) we write

$$h(t) = h_1(t) + h_2(t) = C_o \cos \omega t + D_o \sin \omega t + h_2(t) \tag{2.1o}$$

Then we substitute the expression (2.1o) in (1.1o) which gives

$$U(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^\nu f(t) + \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^\nu h_1 + \tag{2.11}$$

$$+ \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^\nu h_2(t)$$

In this representation, each term on the right-hand side individually satisfies the telegraphic equation. In the same way as we did with the first term of the above solution function (cf.(2.9)) we can also transform the second term:

$$\sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^\nu h_1(t) = (C_o, D_o)\left(\sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} \Omega^\nu \right) \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} \tag{2.12}$$

The first two terms of Eq.(2.11) will be denoted by $U_1(x,t)$, the last one by $v(x,t)$

$$U(x,t) = U_1(x,t) + v(x,t) \quad \text{with} \tag{2.13}$$

$$v(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^\nu h_2(t) \tag{2.14}$$

Now, we determine the coefficients C_o and D_o so that the function $U_1(x,t)$ also satisfies the two boundary conditions (2.3) and (2.4).

One boundary condition, $U_1(o,t)$. follows from (2.11) when we put x=o:

$$U_1(o,t) = f(t) = A_o \cos \omega t + B_o \sin \omega t \qquad (2.15)$$

(cf.(2.3))

The other boundary condition is .

$$U_1(a,t) = o \qquad (2.16)$$

The function v(x,t) is then zero at the ends of the line (x=o and x=a).

Hence, we have another boundary condition for v(x,t), namely

$$v(a,t) = o \qquad (2.17)$$

Moreover, we must determine v(x,t) so that it satisfies the initial conditions (2.1) and (2.2) for $U_1(x,t)$ does not satisfy them.

As we shall see later, the function $U_1(x,t)$ constitutes the steady part of the solution whereas v(x,t) is the non-persistent part of the voltage function.

4. Calculation of the coefficients C_o and D_o

To make the calculation of C_o and D_o more transparent, we bring the matrices

$$P_1 = \sum_{\nu=o}^{\infty} \frac{x^{2\nu}}{(2\nu)!} \Omega^{\nu} \quad \text{and} \quad P_2 = \sum_{\nu=o}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} \Omega^{\nu}$$

to their normal form (cf. /24/, p. 195).

The eigenvalues ϱ_1, ϱ_2 of the matrix

$$\Omega = \begin{pmatrix} \alpha - \gamma\omega^2 & -\beta\omega \\ \beta\omega & \alpha - \gamma\omega^2 \end{pmatrix}$$ are

$$\varrho_{1,2} = (\alpha - \gamma\omega^2) \pm i.\beta\omega \qquad (2.18)$$

The matrix $P_1 = \cosh x\sqrt{\Omega}$ can be brought to diagonal form, because the eigenvalues ϱ_1 and ϱ_2 are different $(\beta \neq 0)$

$$(2.19)\quad T^{-1}\cosh x\sqrt{\Omega}\,T = \sum_{\nu=0}^{\infty}\frac{x^{2\nu}}{(2\nu)!}(T^{-1}\Omega^{\nu}T) = \sum_{\nu=0}^{\infty}\frac{x^{2\nu}}{(2\nu)!}(T^{-1}\Omega\,T)^{\nu} =$$

$$= \sum_{\nu=0}^{\infty}\frac{x^{2\nu}}{(2\nu)!}\Lambda^{\nu} = \sum_{\nu=0}^{\infty}\frac{x^{2\nu}}{(2\nu)!}\begin{pmatrix}\varrho_1^{\nu} & 0\\ 0 & \varrho_2^{\nu}\end{pmatrix} =$$

$$= \begin{pmatrix}\cosh x\sqrt{\varrho_1} & 0\\ 0 & \cosh x\sqrt{\varrho_2}\end{pmatrix}$$

After a simple calculation, we find the matrix

$$(2.20)\quad T = \begin{pmatrix}1 & 1\\ -i & i\end{pmatrix} \quad \text{and its inverse } T^{-1} = \frac{T_{ad}}{|T|}$$

$$(2.21)\quad T^{-1} = \frac{1}{2}\begin{pmatrix}1 & i\\ 1 & -i\end{pmatrix}$$

Multiplying the matrix function (2.19) on the left by T and on the right by T^{-1} we find

$$\cosh x\sqrt{\Omega} = \frac{1}{2}\begin{pmatrix}\cosh x\sqrt{\varrho_1} + \cosh x\sqrt{\varrho_2}, & i(\cosh x\sqrt{\varrho_1} + \cosh x\sqrt{\varrho_2})\\ -i(\cosh x\sqrt{\varrho_1} + \cosh x\sqrt{\varrho_2}), & \cosh x\sqrt{\varrho_1} + \cosh x\sqrt{\varrho_2}\end{pmatrix}$$

After a few transformations

$$\sqrt{\varrho_1} = \sqrt{(\alpha - \gamma\omega^2) - \beta\omega i} = p - iq$$

$$\sqrt{\varrho_2} = \sqrt{(\alpha - \gamma\omega^2) + \beta\omega i} = p + iq$$

with

$$p = \sqrt{\frac{(\alpha - \gamma\omega^2)}{2} + \frac{1}{2}\sqrt{(\alpha - \gamma\omega^2)^2 - (\beta\omega)^2}}$$

$$q = \sqrt{-\frac{(\alpha - \gamma\omega^2)}{2} + \frac{1}{2}\sqrt{(\alpha - \gamma\omega^2)^2 - (\beta\omega)^2}}$$

we find:

$$(2.22)\quad P_1 = \cosh x\sqrt{\Omega} = \begin{pmatrix}p_1 & q_1\\ -q_1 & p_1\end{pmatrix} =$$

$$= \begin{pmatrix}\cosh xp.\cos xq, & \sinh xp.\sin xq\\ -\sinh xp.\sin xq, & \cosh xp.\cos xq\end{pmatrix}$$

In the same way we also treat the matrix

$$(2.23) \qquad P_2 = \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!}\,\Omega^{\nu} = \Omega^{-\frac{1}{2}}.\sinh x\sqrt{\Omega}$$

The calculation gives then

$$(2.24) \qquad P_2 = \begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix} \qquad \text{with}$$

$$p_2 = \frac{1}{p^2+q^2}\,(\,p.\sinh xp.\cos xq + q.\cosh xp.\sin xq\,)$$

$$q_2 = \frac{1}{p^2+q^2}\,(\,p.\cosh xp.\sin xq - q.\sinh xp.\cos xq\,)$$

With x=a, the matrix P_1 becomes

$$M = \sum_{\nu=0}^{\infty} \frac{a^{2\nu}}{(2\nu)!}\,\Omega^{\nu} = \cosh a\sqrt{\Omega} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

with

$$m_{11} = m_{22} = \cosh ap.\cos aq \qquad \text{and}$$

$$m_{12} = -m_{21} = \sinh ap.\sin aq$$

whereas the matrix P_2 becomes

$$N = \sum_{\nu=0}^{\infty} \frac{a^{2\nu+1}}{(2\nu+1)!}\,\Omega^{\nu} = \Omega^{-\frac{1}{2}} \sinh a\sqrt{\Omega} = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix}$$

with

$$n_{11} = n_{22} = \frac{1}{p^2+q^2}\,(p.\sinh ap.\cos aq + q.\cosh ap.\sin aq)$$

$$n_{12} = -n_{21} = \frac{1}{p^2+q^2}\,(p.\cosh ap.\sin aq - q.\sinh ap.\cos aq)$$

Considering the boundary condition (2.14) we obtain

$$U_1(a,t) = (A_o,B_o)\,M\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} + (C_o,D_o)\,N\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} =$$

$$= (A_o m_{11} + B_o m_{21} + C_o n_{11} + D_o n_{21})\cos\omega t +$$

$$+ (A_o m_{12} + B_o m_{22} + C_o n_{12} + D_o n_{22})\sin\omega t = o$$

The coefficients of $\cos\omega t$ and $\sin\omega t$ must vanish in order that the above relation is satisfied at any instant of time t. The

coefficients C_o and D_o are then readily found by Cramer's rule.

$$(2.25)\qquad C_o = \frac{A_o(m_{11}n_{22} - m_{12}n_{21}) + B_o(m_{21}n_{22} - m_{22}n_{21})}{n_{12}n_{21} - n_{11}n_{22}}$$

$$(2.26)\qquad D_o = \frac{A_o(m_{11}n_{12} - m_{12}n_{11}) + B_o(m_{21}n_{12} - m_{22}n_{11})}{n_{11}n_{22} - n_{12}n_{21}}$$

5. Calculation of the voltage part v(x,t)

The condition

$$(2.27)\qquad v(a,t) = \sum_{\nu=0}^{\infty} \frac{a^{2\nu+1}}{(2\nu+1)!} D^{\nu} h_2(t) = 0 \qquad \text{(cf.(2.17))}$$

can be fulfilled when the function of time, $D^{\nu}h_2(t)$, is formally assumed as

$$(2.28)\qquad D^{\nu} h_{2k}(t) = (-1)^{\nu} \cdot h_{2k}(t) \cdot \left[\frac{k^2\pi^2}{a^2}\right]^{\nu} \qquad \text{for } k = 1,2,3,\ldots$$

This relation must be inserted in v(x,t):

$$v_k(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^{\nu} h_{2k}(t) = \sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{a}{k\pi} \frac{x^{2\nu+1}}{(2\nu+1)!} \left(\frac{k\pi}{a}\right)^{2\nu+1} h_{2k}(t)$$

Using the series expansion of the sine and superposing v(x,t) we find

$$(2.29)\qquad v(x,t) = \sum_{k=1}^{\infty} v_k(x,t) = \sum_{k=1}^{\infty} \frac{a}{k\pi} h_{2k}(t) \sin \frac{k\pi x}{a}$$

The Relation (2.28) is equivalent to the equation

$$(2.30)\qquad Dh_{2k}(t) = - h_{2k}(t) \cdot \frac{k^2\pi^2}{a^2} \quad ,$$

whence one can determine the functions h_{2k}.

$$Dh_{2k}(t) = \alpha h_{2k}(t) + \beta \frac{\partial}{\partial t} h_{2k}(t) + \gamma \frac{\partial^2}{\partial t^2} h_{2k}(t) = - h_{2k}(t) \frac{k^2\pi^2}{a^2}$$

$$\gamma h''_{2k}(t) + \beta h'_{2k}(t) + \left(\alpha + \frac{k^2\pi^2}{a^2}\right) h_{2k}(t) = 0$$

This homogeneous differential equation has the solution

$$(2.31) \quad h_{2k}(t) = e^{-\frac{\beta t}{2\gamma}} \left(C_k \cos \omega_k t + D_k \sin \omega_k t\right)$$

where $\omega_k = \sqrt{\left(\alpha + \frac{k^2\pi^2}{a^2}\right)\frac{1}{\gamma} - \frac{\beta^2}{4\gamma^2}}$

(the case $\omega_k^2 < 0$ can easily be inclused in the calculation).

The function $h_2(t)$ can also be obtained by superposing the $h_{2k}(t)$

$$(2.32) \quad h_2(t) = \sum_{k=1}^{\infty} h_{2k}(t)$$

Inserting (2.32) into (2.29) we obtain

$$(2.33) \quad v(x,t) = e^{-\frac{\beta t}{2\gamma}} \sum_{k=1}^{\infty} \frac{a}{k\pi} \left(C_k \cos \omega_k t + D_k \sin \omega_k t\right) \sin \frac{k\pi x}{a}$$

6. The solution U(x,t)

Inserting (2.33) into (2.13) we obtain the solution function

$$(2.34) \quad U(x,t) = (A_0, B_0)\begin{pmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{pmatrix}\begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} + (C_0, D_0)\begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix}\begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix}$$
$$+ e^{-\frac{\beta t}{2\gamma}} \sum_{k=1}^{\infty} \frac{a}{k\pi} \left(C_k \cos \omega_k t + D_k \sin \omega_k t\right) \sin \frac{k\pi x}{a}$$

The coefficients C_k and D_k follow from the initial conditions (2.1) and (2.2).

We insert the condition (2.1) in (2.34) and we obtain:

$$U(x,0) = (A_0, B_0)\begin{pmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + (C_0, D_0)\begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} +$$
$$+ \sum_{k=1}^{\infty} \frac{a}{k\pi} C_k \sin \frac{k\pi x}{a} = 0$$

The $\frac{a}{k\pi} C_k$ can be regarded as the Fourier coefficients of the orthogonal system $\left\{ \sin \frac{k\pi x}{a} \right\}$, for $\sum_{k=1}^{\infty} \frac{a}{k\pi} C_k \sin \frac{k\pi x}{a}$ represents a Fourier series for the function

$$\left\{ - (A_o, B_o) \begin{pmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{pmatrix} \begin{pmatrix} 1 \\ o \end{pmatrix} - (C_o, D_o) \begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix} \begin{pmatrix} 1 \\ o \end{pmatrix} \right.$$

Hence, we obtain the Fourier coefficients

$$(2.35) \quad \frac{a}{k\pi} C_k = - \frac{2}{a} \int_o^a \left\{ (A_o, B_o) \begin{pmatrix} p_1 \\ -q_1 \end{pmatrix} + (C_o, D_o) \begin{pmatrix} p_2 \\ -q_2 \end{pmatrix} \right\} \sin \frac{k\pi x}{a} \, dx$$

The result for the coefficients C_k is

$$(2.36) \quad C_k = - \frac{2k\pi}{a^2} \int_o^a U_1(x,o) \sin \frac{k\pi x}{a} \cdot dx$$

To determine the coefficients D_k we have to consider the initial condition (2.2).

$$U_t(x,o) = U_{1t}(x,o) + v_t(x,o) = o$$

Calculation gives

$$U_t(x,o) = (A_o, B_o) \begin{pmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{pmatrix} \begin{pmatrix} c \\ 1 \end{pmatrix} \omega + (C_o, D_o) \begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix} \begin{pmatrix} o \\ 1 \end{pmatrix} \omega -$$

$$- \frac{\beta}{2\gamma} \sum_{k=1}^{\infty} (C_k, D_k) \frac{a}{k\pi} \begin{pmatrix} 1 \\ o \end{pmatrix} \sin \frac{k\pi x}{a} +$$

$$+ \sum_{k=1}^{\infty} (C_k, D_k) \frac{a}{k\pi} \begin{pmatrix} o \\ 1 \end{pmatrix} \omega_k \sin \frac{k\pi x}{a} = o$$

Here, $\sum_{k=1}^{\infty} D_k \frac{a}{k\pi} \omega_k \sin \frac{k\pi x}{a}$ can again be assumed to be the Fourier series of the function

$$\left\{ - (A_o, B_o) \begin{pmatrix} q_1 \\ p_1 \end{pmatrix} \cdot \omega - (C_o, D_o) \begin{pmatrix} q_2 \\ p_2 \end{pmatrix} \omega + \right.$$

$$\left. + \frac{\beta}{2\gamma} \left[-(A_o, B_o) \begin{pmatrix} p_1 \\ -q_1 \end{pmatrix} - (C_o, D_o) \begin{pmatrix} p_2 \\ -q_2 \end{pmatrix} \right] \right\}$$

with the Fourier coefficients

$$(2.37) \quad \frac{a}{k\pi} D_k \omega_k = \frac{2}{a} \int_o^a \left\{ - U_{1t}(x,o) - \frac{\beta}{2\gamma} U_1(x,o) \right\} \sin \frac{k\pi x}{a} \cdot dx$$

7. The solution J(x,t)

Integrating (1.1) we obtain the solution for the current J(x,t)

$$(2.39) \qquad J(x,t) = -\int \left(\varrho + c\,\frac{\partial}{\partial t}\right) U(x,t)\,dx + J_o(t)$$

With the abbreviations

$$P_1(x) = \begin{pmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{pmatrix} = \cosh x\sqrt{\Omega}$$

$$P_2(x) = \begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix} = \Omega^{-\frac{1}{2}} \sinh x\sqrt{\Omega}$$

$$P_3(x) = \int P_2(x)\,dx = \Omega^{-1} \cosh x\sqrt{\Omega}$$

we find:

$$(2.40) \quad J(x,t) = -\varrho\left\{(A_o,B_o)\,P_2\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} + (C_o,D_o)\,P_3\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} - \right.$$

$$\left. - e^{-\frac{\beta t}{2\gamma}} \sum_{k=1}^{\infty}\left(\frac{a}{k\pi}\right)^2\left[C_k\cos\omega_k t + D_k\sin\omega_k t\right]\cos\frac{k\pi x}{a}\right\} -$$

$$- c\left\{(A_o,B_o)\,P_2\begin{pmatrix}-\sin\omega t\\ \cos\omega t\end{pmatrix}\omega + (C_o,D_o)\,P_3\begin{pmatrix}-\sin\omega t\\ \cos\omega t\end{pmatrix}\omega + \right.$$

$$+ \frac{\beta}{2\gamma} e^{-\frac{\beta t}{2\gamma}} \sum_{k=1}^{\infty}\left(\frac{a}{k\pi}\right)^2\left[C_k\cos\omega_k t + D_k\sin\omega_k t\right]\cos\frac{k\pi x}{a} -$$

$$\left. - e^{-\frac{\beta t}{2\gamma}} \sum_{k=1}^{\infty}\left(\frac{a}{k\pi}\right)^2\left[-C_k\omega_k\sin\omega_k t + D_k\omega_k\cos\omega_k t\right]\cos\frac{k\pi x}{a}\right\} + J_o(t)$$

The integration constant $J_o(t)$ can be found from the relation for J(o,t) which can be derived directly from the basic equation (1.1).

The time function J(o,t) satisfies the differential equation (1.1)

$$(1.1) \qquad U_x(o,t) = h(t) = -\left(r + l\,\frac{\partial}{\partial t}\right) J(o,t)$$

$$(2.41) \qquad l J_t(o,t) + r J(o,t) = h(t)$$

Its solution is

$$(2.42)\quad J(o,t) = -\frac{C_o l}{r^2+\omega^2 l^2}\left(\frac{r}{l}\cos\omega t + \omega\sin\omega t\right) -$$

$$-\frac{D_o l}{r^2+\omega^2 l^2}\left(\frac{r}{l}\sin\omega t - \omega\cos\omega t\right) -$$

$$-\sum_{k=1}^{\infty}\frac{C_k}{l}\frac{e^{-\frac{\beta t}{2\gamma}}}{\left(\frac{r}{l}-\frac{\beta}{2\gamma}\right)^2+\omega_k^2}\left[\left(\frac{r}{l}-\frac{\beta}{2\gamma}\right)\cos\omega_k t + \omega_k\sin\omega_k t\right] -$$

$$-\sum_{k=1}^{\infty}\frac{D_k}{l}\frac{e^{-\frac{\beta t}{2\gamma}}}{\left(\frac{r}{l}-\frac{\beta}{2\gamma}\right)^2+\omega_k^2}\left[\left(\frac{r}{l}-\frac{\beta}{2\gamma}\right)\sin\omega_k t - \omega_k\cos\omega_k t\right] + Ke^{-\frac{rt}{l}}$$

The constant K results from the initial condition $J(x,o) = 0$ (cf.(2.2))

$$(2.43)\qquad J(o,o) = o$$

$$(2.44)\quad K = \frac{C_o r - D_o l\omega}{r^2+\omega^2 l^2} + \sum_{k=1}^{\infty}\frac{C_k\left(\frac{r}{l}-\frac{\beta}{2\gamma}\right) - D_k\omega_k}{l\left[\left(\frac{r}{l}-\frac{\beta}{2\gamma}\right)^2+\omega_k^2\right]}$$

Hence, we find the original integration constant $J_o(t)$

$$(2.45)\quad J_o(t) = J(o,t) + c\left\{(C_o,D_o)\,\Omega^{-1}\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} -\right.$$

$$\left. - e^{-\frac{\beta t}{2\gamma}}\sum_{k=1}^{\infty}\left(\frac{a}{k\pi}\right)^2\left[C_k\cos\omega_k t + D_k\sin\omega_k t\right]\right\} +$$

$$+ c\left\{(C_o,D_o)\,\Omega^{-1}\begin{pmatrix}-\sin\omega t\\ \cos\omega t\end{pmatrix}\omega \quad +\right.$$

$$+\frac{\beta}{2\gamma}e^{-\frac{\beta t}{2\gamma}}\sum_{k=1}^{\infty}\left(\frac{a}{k\pi}\right)^2\left[C_k\cos\omega_k t + D_k\sin\omega_k t\right] -$$

$$\left. - e^{-\frac{\beta t}{2\gamma}}\sum_{k=1}^{\infty}\left(-C_k\omega_k\sin\omega_k t + D_k\omega_k\cos\omega_k t\right)\right\}$$

8. Numerical examples

Two numerical examples shall illustrate the switch-on transients problem in a shorted line. We have examined the maximum values of the power function at the point x=o in dependence on the phase angle τ of the applied voltage

$$U(o,t) = u \cos(\omega t - \tau) = A_o \cos \omega t + B_o \sin \omega t$$

during the transient process. Calculations were performed at the ZUSE Z23 computer of Innsbruck University.+)

In the first numerical example, the electrical constants per unit lenght were chosen as follows:

resistance	r = 1	lenght of wires	a = 1
induktive reactance	l = o,2	angular frequency of voltage	ω = 3oo
capacitive reactance	c = o,oo2		
leakage	g = o		

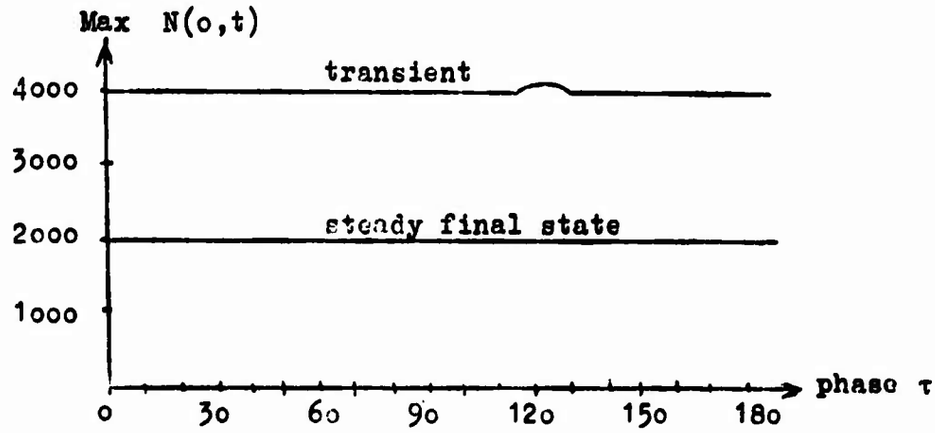
This result (see table) shows that there is a notable strong resonance and superposition effekt. For $\tau = 12o^{o}$, the power funotion main peaks at x=o increases to more than 2o5% of the maximum value in the steady final state.

+) The corresponding programs of the numerical examples are contained in(/43/p. 39 ff.)

The following table shows the results:

τ	$A_o = 100.\cos\tau$	$B_o = 100.\sin\tau$	Main peaks of power function during transient	Peaks in final state
0	100	0	3985	1992
10	98,4800772	17,3648176	3995	1992
20	93,9692619	34,2020142	4003	1992
30	86,6025401	50,0000000	4004	1992
40	76,6044441	64,2787608	4015	1992
50	64,2787608	76,6044441	4018	1992
60	50,0000000	86,6025401	4021	1992
70	34,2020142	93,9692619	4026	1992
80	17,3648176	98,4800772	4028	1992
90	0	100	4030	1992
100	-17,3648176	98,4800772	4032	1992
110	-34,2020142	93,9692619	4025	1992
120	-50,0000000	86,6025401	4090	1992
130	-64,2787608	76,6044441	4004	1992
140	-76,6044441	64,2787608	4040	1992
150	-86,6025401	50,0000000	4009	1992
160	-93,9692619	34,2020142	4005	1992
170	-98,4800772	17,3648176	4002	1992
180	-100	0	3985	1992
190	-98,4800772	-17,3648176	3995	1992
200	-93,9692619	-34,2020142	4003	1992
210	-86,6025401	-50,0000000	4004	1992

Fig. Peaks of the power function N(o,t) = J(o,t).U(o,t) versus phase angle τ of the applied alternating voltage.



The following values were chosen in the second numerical example:

resistance	r = 1	lenght of wires	a = 1
inductive reactance	l = 1	angular frequency of voltage	ω = 1oo
capacitive reactance	c = o,o1		
leakage	ϱ = o,o1		

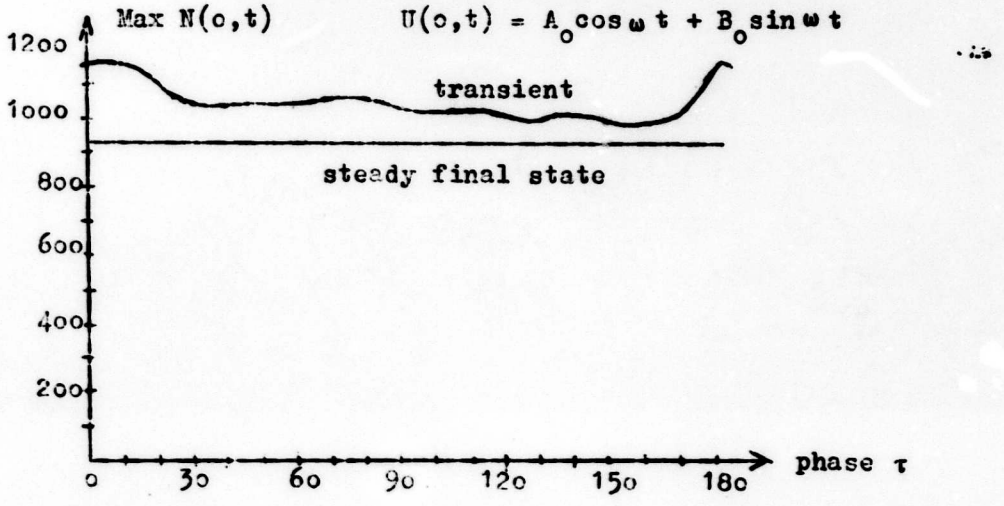
With N(o,t) tabulated, the following result was obtained for the peaks of the power function N(x,t) at the point x=o

Phase τ	A_o=1oo.cos τ	B_o=1oo.sin τ	Main peaks during transient	Peaks in steady final state
o	1oo	o	1144	928
1o	98,48oo772	17,3648176	1166	928
2o	93,9692619	34,2o2o142	1127	928
3o	86,6o254o1	5o,oooooo	1o28	928
4o	76,6o44441	64,2787608	1o29	928
5o	64,2787608	76,6o44441	1o37	928
6o	5o,oooooo	86,6o254o1	1o4o	928
7o	34,2o2o142	93,9692619	1o42	928
8o	17,3648176	98,48oo772	1o68	928

Phase τ	A_o=1oo.cos τ	B_o=1oo.sin τ	Main peaks during transient	Peaks in steady final state
8o	17,3648176	98,48oo772	1o68	928
9o	o	1oo	1o43	928
1oo	-17,3648176	98,48oo772	1o42	928
11o	-34,2o2o142	93,9692619	1o38	928
12o	-5o,oooooo	86,6o254o1	1o36	928
13o	-64,2787608	76,6o44441	1o12	928
14o	-76,6o44441	64,2787608	1o24	928
15o	-86,6o254o1	5o,oooooo	1o27	928
16o	-93,9692619	34,2o2o142	1oo8	928
17o	-98,48oo772	17,3648176	1o36	928
18o	-1oo	o	1144	928
19o	-98,48oo772	-17,3648176	1166	928
2oo	-93,9692619	-34,2o2o142	1127	928

This example shows a clear dependence of the main peaks on the phase angle τ of the applied alternating voltage.

Fig. Main peaks of the power function N(o,t) during the transient and peaks of N(o,t) in the steady final state versus phase angle τ of the applied alternating voltage



VII.3 Switch-on Transients with Open Wires

The transient processes with the wire ends open can be treated mathematically analogous to the case of shorted ends.

1. Initial and boundary conditions

We shall assume the initial conditions

(3.1) $U(x,o) = o \qquad x > o$

(3.2) $J(x,o) = o$

(3.3) $J_t(x,o) = o$

and the boundary conditions

(3.4) $U(o,t) = A \cos \omega t + B \sin \omega t$

(3.5) $J(a,t) = o$

Like in (1.1o), the formal solution for the current is

(3.6) $$J(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^{\nu} \overline{f}(t) + \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^{\nu} \overline{h}(t)$$

where

(3.7) $\overline{f}(t) = J(o,t)$

and

(3.8) $\overline{h}(t) = J_x(o,t)$

The function $\overline{h}(t)$ can be found from the basic equation (1.1) and from (3.4)

(3.9) $$\overline{h}(t) = J_x(o,t) = - \varrho U(o,t) - c U_t(o,t) =$$
$$= \overline{A} \cos \omega t + \overline{B} \sin \omega t$$

where we have put

$$-B\omega c - \varrho A = \overline{A} \qquad \text{and}$$
$$A\omega c - \varrho B = \overline{B}$$

2. Statement for J(o,t)

For the function J(o,t) we set formally (cf. VII.2.3)

$$(3.10)\quad J(o,t) = \overline{f}(t) = \overline{f}_1(t) + \overline{f}_2(t) = \overline{C}_o \cos\omega t + \overline{D}_o \sin\omega t + \overline{f}_2(t)$$

which we insert into (3.6)

$$(3.11)\quad J(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^{\nu} \overline{f}_1(t) + \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^{\nu} \overline{f}_2(t) + \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} D^{\nu} \overline{h}(t)$$

We call the sum of the first and third terms $J_1(x,t)$ and transform them in the same way as in (2.9)

$$(3.12)\quad J_1(x,t) = (\overline{C}_o, \overline{D}_o)\left(\sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} \Omega^{\nu}\right)\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} + (\overline{A}_o, \overline{B}_o)\left(\sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu+1)!} \Omega^{\nu}\right)\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} = (\overline{C}_o, \overline{D}_o)\, P_1\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} + (\overline{A}_o, \overline{B}_o)\, P_2\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix}$$

... the second term in (3.11) we write

$$(3.13)\quad w(x,t) = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!} D^{\nu} \overline{f}_2(t)$$

$$(3.11')\quad J(x,t) = J_1(x,t) + w(x,t)$$

The coefficients $\overline{C}_o$ and $\overline{D}_o$ must be determined so that the function $J_1(x,t)$ satisfies the boundary condition (3.5)

$$(3.14)\quad J_1(a,t) = o$$

$J_1(x,t)$ does not satisfy the initial conditions.

The function w(x,t) must be determined so that the initial conditions are fulfilled and that w(x,t) becomes zero at the end of the line (x=a).

Thus, we have further boundary condition for $w(x,t)$

(3.15) $\quad w(a,t) = 0$

3. Calculation of coefficients $\overline{C}_o$, $\overline{D}_o$ and of the function $w(x,t)$

Using the boundary conditions (3.14) we obtain (cf. VII.2.4)

$$(3.16)\quad J_1(a,t) = (\overline{C}_o,\overline{D}_o)\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22}\end{pmatrix}\begin{pmatrix}\cos\omega t \\ \sin\omega t\end{pmatrix} + \\ + (\overline{A}_o,\overline{B}_o)\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22}\end{pmatrix}\begin{pmatrix}\cos\omega t \\ \sin\omega t\end{pmatrix} = o$$

This relation holds at any time t. Therefore, the coefficients of $\sin\omega t$ and $\cos\omega t$ must be zero. The coefficients A and B are given by the boundary condition (3.4), whence the coefficients $\overline{C}_o$ and $\overline{D}_o$ can be found by means of Cramer's rule; the result is

$$(3.17)\quad \overline{C}_o = \frac{\overline{A}_o(n_{11}m_{22} - n_{12}m_{21}) + \overline{B}_o(n_{21}m_{22} - n_{22}m_{21})}{m_{12}m_{21} - m_{11}m_{22}}$$

$$\overline{D}_o = \frac{\overline{A}_o(n_{11}m_{12} - n_{12}m_{11}) + \overline{B}_o(n_{21}m_{12} - n_{22}n_{11})}{m_{22}m_{11} - m_{21}m_{12}}$$

The condition (3.15)

$$w(a,t) = \sum_{\nu=0}^{\infty} \frac{a^{2\nu}}{(2\nu)!}\, D^{\nu}\overline{f}_2(t) = o$$

is satisfied if we put

$$(3.18)\quad D^{\nu}\overline{f}_2(t) = (-1)^{\nu}\left[\left(\frac{2k+1}{2a}\right)\pi\right]^{2\nu}\overline{f}_{2k} \qquad \text{for } k = o,1,2,\ldots.$$

$$w_k(t) = \sum_{\nu=0}^{\infty} (-1)^{\nu}\,\frac{x^{2\nu}}{(2\nu)!}\left(\frac{2k+1}{2a}\pi\right)^{2\nu}\cdot \overline{f}_{2k}(t) =$$

$$= \cos\left(\frac{2k+1}{2a}\,\pi x\right)\overline{f}_{2k}(t)$$

Superposing the $w_k(x,t)$ we find the current part $w(x,t)$

$$(3.19) \qquad w(x,t) = \sum_{k=0}^{\infty} \bar{f}_{2k}(t) \cos \frac{(2k+1)}{2a}\pi x$$

The relation (3.18) can be written as a differential equation of the form

$$(3.20) \quad D\bar{f}_{2k}(t) = \alpha \bar{f}_{2k}(t) \beta \frac{\partial}{\partial t} \bar{f}_{2k}(t) + \gamma \frac{\partial^2}{\partial t^2} \bar{f}_{2k}(t) =$$

$$= (-1).\bar{f}_{2k}\left(\frac{2k+1}{2a}\pi\right)^2$$

$$\gamma \bar{f}''_{2k}(t) + \beta \bar{f}'_{2k}(t) + \left[\alpha + \left(\frac{2k+1}{2a}\pi\right)^2\right] \bar{f}_{2k}(t) = 0$$

The functions $\bar{f}_{2k}(t)$ can then be found from this equation.

$$(3.21) \quad \bar{f}_{2k}(t) = e^{-\frac{\beta t}{2\gamma}} (\bar{C}_k \cos \omega_k t + \bar{D}_k \sin \omega_k t)$$

with

$$\omega_k = \sqrt{\left[\alpha + \left(\frac{2k+1}{2a}\pi\right)^2\right] \frac{1}{\gamma} - \frac{\beta^2}{4\gamma^2}}$$

where we have assumed that

$$\frac{1}{\gamma}\left(\alpha + \left(\frac{\pi}{2a}\right)^2\right) > \frac{\beta^2}{4\gamma^2}$$

4. The solution $J(x,t)$ and $U(x,t)$

Inserting (3.21) in (3.19) we obtain

$$(3.22) \quad J(x,t) = J_1(x,t) + w(x,t) =$$

$$= (\bar{C}_0, \bar{D}_0)\begin{pmatrix} p_1 & q_1 \\ -q_1 & p_1 \end{pmatrix}\begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} + (\bar{A}_0, \bar{B}_0)\begin{pmatrix} p_2 & q_2 \\ -q_2 & p_2 \end{pmatrix}\begin{pmatrix} \cos \omega \\ \sin \end{pmatrix}$$

$$+ e^{-\frac{\beta t}{2\gamma}} \sum_{k=0}^{\infty} (\bar{C}_k \cos \omega_k t + \bar{D}_k \sin \omega_k t) \cos \frac{(2k+1)}{2a} \pi x$$

The coefficients $\bar{C}_k$ and $\bar{D}_k$ can be calculated by the same reasoning as in VII.2.6.

Considering the initial conditions (3.2) and (3.3) we obtain the result

$$(3.23)\quad \bar{C}_k = -\frac{2}{a}\int_0^a J_1(x,o)\cos\frac{(2k+1)\pi x}{2a}\,dx$$

$$(3.24)\quad \bar{D}_k = -\frac{2}{a\omega_k}\int_0^a\left\{J_{1\ t}(x,o) + \frac{\beta}{2\gamma}J_1(x,o)\right\}\cos\frac{(2k+1)\pi x}{2a}$$

The solution for the voltage U(x,t) is found by integration of (1,2):

$$(3.25)\quad U(x,t) = -\int\left(rJ + l\frac{\partial J}{\partial t}\right)dx + U_o(t) =$$

$$= -r\left\{(\bar{C}_o,\bar{D}_o)\,P_2\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} + (\bar{A}_o,\bar{B}_o)\,P_3\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix} + \right.$$

$$\left. + e^{-\frac{\beta t}{2\gamma}}\sum_{k=o}^{\infty}\frac{2a}{(2k+1)\pi}(\bar{C}_k\cos\omega_k t + \bar{D}_k\sin\omega_k t)\sin\frac{(2k+1)\pi x}{2a}\right\} -$$

$$- l\left\{(\bar{C}_o,\bar{D}_o)\,P_2\begin{pmatrix}-\sin\omega t\\ \cos\omega t\end{pmatrix}\omega + (\bar{A}_o,\bar{B}_o)\,P_3\begin{pmatrix}-\sin\omega t\\ \cos\omega t\end{pmatrix}\omega\right. -$$

$$- \frac{\beta}{2\gamma}e^{-\frac{\beta t}{2\gamma}}\sum_{k=o}^{\infty}(\bar{C}_k\cos\omega_k t + \bar{D}_k\sin\omega_k t)\frac{2a}{(2k+1)\pi}\sin\frac{(2k+1)\pi x}{2a} +$$

$$+ e^{-\frac{\beta t}{2\gamma}}\sum_{k=o}^{\infty}\frac{2a}{(2k+1)\pi}(-\bar{C}_k\omega_k\sin\omega_k t + \bar{D}_k\omega_k\cos\omega_k t)\cdot$$

$$\left.\cdot\sin\frac{(2k+1)\pi x}{2a}\right\} + U_o(t)$$

where

$$U_o(t) = U(o,t) + r\left\{\diagup\right\} + l\left\{\diagdown\right\} =$$

$$= A\cos\omega t + B\sin\omega t +$$

$$+ r\left\{(\bar{A}_o,\bar{B}_o)\,\Omega^{-1}\begin{pmatrix}\cos\omega t\\ \sin\omega t\end{pmatrix}\right\} + l\left\{(\bar{A}_o,\bar{B}_o)\,\Omega^{-1}\begin{pmatrix}-\sin\omega t\\ \cos\omega t\end{pmatrix}\omega\right\}$$

REFERENCES

1. J.C.BUTCHER, Coefficients for the study of Runge-Kutta integration processes, *J. Austral. Math. Soc. 3*, (1963), 185-201.

2. J.C.BUTCHER, Implicit Runge-Kutta processes, *Math.Comp. 18*, (1964), 50-64.

3. F.CAP, J.MENNIG, Analytical method for determining n-group neutron fluxes in cylindrical shielding problems using Lie series, *Nucleonic 6*, (1964), 141-147.

4. K.T.CHEN, On a generalization of Picard's approximation, *J. Differential Eqs. 2*, (1966), 438-448.

5. C.CHIARELLA, A.REICHEL, On the evaluation of Integrals related to the error function, *Math. Comp. 22*, (1968), 137-143.

6. E.A.CODDINGTON, N.LEVINSON, Theory of ordinary differential equations, Mc. Graw-Hill 1955.

7. L.COLLATZ, The numerical treatment of differential equations, Spronger Verlag 1960.

8. G.J.COOPER, Interpolation and quadrature methods for ordinary differential equations, *Math.Comp.22*,(1968), 69-76.

9. A.DEPRIT, R.V.M.ZAHAR, Numerical integration of an orbit and its concomitant variations by recurrent power series, ZAMP 17 (1966), 425-430.

10. H.L.Durham et al., Study of methods for the numerical solution of ordinary differential equations. Final Report, Project A-740, NASA N 65-20106, 1964.

11. F.ERWE, Gewöhnliche Differentialgleichungen, B.I. Htb.19, Mannheim 1961.

12. E.FEHLBERG, New high-order Runge-Kutta formulas ..., ZAMM 44 (1964), T17-T29.

13. E.FEHLBERG, New high order Runge-Kutta formulas with an arbitrary small truncation error, ZAMM 46 (1966),1-16.

14. E.FEHLBERG, Zur numerischen Integration von Differentialgleichungen durch Potenzreihenansätze..., ZAMM 44 (1964), 83-88.

15. S.FILIPPI, Neue Gauss-Typ Quadraturformeln, Habilitationsschrift, Th. Aachen 1964.

16. W.GAUTSCHI, Computation of successive derivatives of f(z)/z, Math.Comp. 20 (1966), 209-214.

17. W.GAUTSCHI, On inverses of Vandermonde and confluent Vandermunde matrices I, II, Num. Math. 4 (1962),117-123 and Num.Math. 5 (1963), 423-430.

18. A.GIBBONS, A program for the automatic integration of diff. eq. using the method of Taylor series, Computer J. 3 (1960), 108-111.

19. W.GLASMACHER, D.SOMMER, Implizite Runge-Kutta-Formeln, Forschungsber. d. Landes Nordrh.Westf. 1763, 1966.

20. H.GREENSPAN, W.HAFNER, M.RIBARIC, On varying step size in numerical int. of first order diff. eq., Num.Math. 7 ((1965), 286-291).

21. W.GRÖBNER, Die Lie-Reihen und ihre Anwendungen, D. Verlag der Wiss. Berlin 1960, 1967.

22. W.GRÖBNER , H.KNAPP eds., Contributions to the method of Lie series, B.I.Htb. 802/802a, Mannheim 1967.

23. W.GRÖBNER, P.LESKY, Mathematische Methoden der Physik II, B.I.Htb 90/90a.

24. W.GRÖBNER, Matrizenrechnung, B.I.Htb 1o3/1o3a

25. K.H. KASTLUNGER, Runge-Kutta-Formeln mit mehrfachen Knoten, Thesis, Univ. Innsbruck 1969

26. H.KNAPP, Über eine Verallgemeinerung des Verfahrens der sukzessiven Approximation zur Lösung von Differentialgleichungssystemen. Mh.Math.68 (1964), 33-45.

27 H.KNAPP, G.WANNER, On the numerical treatment of ordinary differential equations, Ch. II in W.GRÖBNER, H.KNAPP eds , |22|., 1967

28. H.KNAPP, G.WANNER, Numerische Integration gewöhnlicher Differentialgleichungen; Einschrittverfahren, in D.LAUGWITZ ed , Überblicke Mathematik, B.I.Htb.161, Mannheim 1968, 87-114

29 H.KNAPP, G.WANNER, Num sol. of ord diff.eq. by Gröbner's method of Lie series, MRC Techn.Sum.Rep. 880, Math. Research Center, Madison Wisconsin, 1968

30. H.KNAPP, G.WANNER, Liese, a program for ord.diff.eq using Lie series, MRC Techn.Rep. 881, MRC, Univ.Wisconsin, Madison 1968

31. V.I.KRYLOV, V.V.LUGIN, L.A.JANOVICH, Tafeln für die num Integration ... $\int_0^1 x^b(1-x)^a f(x)dx$, Minsk 1963

32. J.A.LEAVITT, Methods and applications of power series, Math.Comp. 20 (1966), 46-52

33 G.MAESS, Quantitative Verfahren zur Bestimmung per. Lösungen aut.nichtl Diffgl., Abh.Dtsch.Akad.Wiss. Kl Math., Heft 3 (1965).

34. G.MAESS, Zur Bestimmung der Restgliedes von Lie-Reihen, Wiss Z.Friedrich-Schiller-Univ. Jena, Math.natur. Reihe 14 (1965), 423-425.

35. J.MILLER, R.P.HURST, Simplified calculation of the exp. integral, MTAC 12 (1958), 187-193

36. R.E.MOORE, Interval analysis, Prentice Hall 1966.

37. D.MORRISON, Optimal mesh size in num int.of ord.diff.eq., J.Assos.f.Comput.Mach.,9 (1962), 98-103.

38 I.P.NATANSON, Konstruktive Funktionentheorie, Akad. Verlag Berlin.

39. E.RABE, Detrmin. ... periodic trojan orbits..., Astron.J. 66 (1961), 500-513.

40. L.B.RALL, ed., Error in digital computation I, Wiley New York 1965

41. R.D.RICHTMYER, Detached-shock calculations by power series. I, A.E.C.Research Report, NYU-7973, Courant Inst. of Math.Sci. 1957.

42. J.B.ROSSER, A Runge-Kutta for all seasons, SIAM Rev 9, (1967), 417-452.

43. R.SÄLY, Berechnung des Einschaltvorganges der Telegraphengl. mit Hilfe von verallg. Lie-Reihen, Thesis Innsbruck, 1969.

44. D.SHANKS, Math.Comp. 18 (1964), 75-86.

45. D.SOMMER, Neue impl. Runge-Kutta Formeln..., Diss.Aachen 1967.

46. J.F.STEFFENSEN, K.danske Vidensk. Selsk., Mat.fys.Medd. 30 (1956) No.18

47. A.J.STRECOK, Math.Comp. 22 (1968), 144-158.

48. D.D.STancu, A.H.STROUD, Quadrature formulas with simle Gaussian and multiple fixed nodes, Math.Comp. 17 (1963), 384-394.

49. A.H.STROUD, D.SECREST, Gaussian quadrature formulas, Prentice Hall 1966.

50. A.H.STROUD, D.D.STANCU, Quadrature f. with multiple gaussian nodes J.SIAM Nume.Ana. Ser.B 2 (1965),129-143.

51. G.WANNER, Integration gewöhnlicher Differentialgleichungen, B.I.Htb 831, Mannheim 1969.

52. W.WASOW, Asymptotic expansions for ord.diff.eqs., Interscience 1965.

53 G.MARGREITER. Über die Gröbnersche Methode der Lie-Reihen für gewöhnliche Diffgl., Thesis, Univ.Innsbruck 1969.

54. V.M.ALEKSEEV, Vestnik Moskov.Univ.Ser 1 Mat.Meh 1961 no 2.

55. F.BRAUER, The asymptotic behaviour of perturbed nonl. systems, in A.GHIZZETTI ed., Stability problems of solutionsof diff.eqs., Gubbio 1966, p.51-56.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Prof. W. Gröbner Dept. of Mathematics Univ. of Innsbruck, Austria	UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

Development of New Methods for the Solution of Differential Equations by the Method of Lie Series

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Technical Report June 1968 - July 1969

5. AUTHOR(S) (First name, middle initial, last name)

W. Gröbner, K. H. Kastlunger, H. Reitberger, R. Saly, G. Wanner.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July 1969	156	55

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	US Army Research & Development Group (Europe) APO New York 09757

13. ABSTRACT

This report summarizes the recent work in the application of the LIE-series method to the solution of ordinary and partial differential equations.
After a short introduction the power series method which is a special case of the Lie series method is described. Further discussed is the interesting concept of recursion formulas and the calculation of the "transfer matrix"(connection matrix), the derivatives of the solution with respect to the initial values.
The next portion of the report deals with the numerical evaluation of the Lie series perturbation formula. This portion contains the results of the report /29/, which has been written together with H. Knapp at the MRC, Madison, Wisconsin. Suitable quadrature formulas and recursions, statements on the order and error estimation are given. Numerical examples finish the chapter and compare the method also with that of Fehlberg.
Gröbner's integral equation is proved which leads to short proofs and to various generalizations of the method.
The concept of Runge-Kutta is generalized to methods with multiple nodes, which is possible with the use of the Lie differential operator D. A general theory is developed and the method of Fehlberg is shown to be a special case.
Step-size control and the application of generalized Lie series to the calculation of switch-on transients occuring in the telegraphic equation is discussed.

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Solution of Differential Equations Lie Series Power Series Runge-Kutta methods						

UNCLASSIFIED
Security Classification